

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:47:27 ; Search time 1334.93 Seconds  
(without alignments)  
1927.840 Million cell updates/sec

Title: US-09-214-679-1  
Perfect score: 1442  
Sequence: 1 cccgggaactccatgtggcc.....aatgcaattcattggatcc 1442

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 972840 seqs, 892348106 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_hum1:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_vi:\*  
31: gb\_htg1:\*  
32: gb\_htg2:\*  
33: gb\_in1:\*  
34: gb\_in2:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pr4:\*  
40: gb\_htg3:\*  
41: gb\_htg4:\*  
42: gb\_htg5:\*  
43: gb\_htg6:\*

44: gb\_htg7:\*  
45: em\_htg1:\*  
46: em\_htg2:\*  
47: em\_htg3:\*  
48: em\_hum5:\*  
49: gb\_pl3:\*  
50: gb\_pr5:\*  
51: gb\_htg8:\*  
52: gb\_htg9:\*  
53: gb\_htg10:\*  
54: gb\_htg11:\*  
55: gb\_htg12:\*  
56: gb\_htg13:\*  
57: gb\_htg14:\*  
58: gb\_in3:\*  
59: gb\_htg15:\*  
60: gb\_htg16:\*  
61: gb\_htg17:\*  
62: em\_htg4:\*  
63: em\_htg5:\*  
64: em\_htg6:\*  
65: em\_htg7:\*  
66: em\_hum6:\*  
67: gb\_htg18:\*  
68: gb\_htg19:\*  
69: gb\_htg20:\*  
70: gb\_htg21:\*  
71: gb\_htg22:\*  
72: gb\_htg23:\*  
73: gb\_htg24:\*  
74: gb\_htg25:\*  
75: gb\_htg26:\*  
76: gb\_htg27:\*  
77: gb\_htg28:\*  
78: gb\_htg29:\*  
79: gb\_htg30:\*  
80: gb\_htg31:\*  
81: gb\_vil:\*  
82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Query |       |        | DB | ID         | Description          |
|------------|-------|-------|--------|----|------------|----------------------|
|            | Score | Match | Length |    |            |                      |
| 1          | 1442  | 100.0 | 1442   | 5  | A69475     | A69475 Sequence 1    |
| 2          | 1442  | 100.0 | 1442   | 5  | A72152     | A72152 Sequence 1    |
| C 3        | 20    | 1.4   | 1100   | 77 | AC057459   | AC057459 Giardia     |
| C 4        | 20    | 1.4   | 3167   | 2  | RMU83439   | RMU83439 Rickettsia  |
| C 5        | 20    | 1.4   | 3176   | 2  | AF018076   | AF018076 Rickettsia  |
| C 6        | 20    | 1.4   | 3176   | 2  | RSU83455   | RSU83455 Rickettsia  |
| C 7        | 20    | 1.4   | 3176   | 2  | TTU83456   | TTU83456 Thai tick t |
| C 8        | 20    | 1.4   | 3179   | 2  | RMU83447   | RMU83447 Rickettsia  |
| C 9        | 20    | 1.4   | 3182   | 2  | AF120018S2 | AF120018 Rickettsia  |
| C 10       | 20    | 1.4   | 3182   | 2  | AF120020S1 | AF120020 Rickettsia  |
| C 11       | 20    | 1.4   | 3182   | 2  | AF120022S2 | AF120022 Rickettsia  |
| C 12       | 20    | 1.4   | 3182   | 2  | RAU83436   | RAU83436 Rickettsia  |
| C 13       | 20    | 1.4   | 3182   | 2  | RCU83440   | RCU83440 Rickettsia  |
| C 14       | 20    | 1.4   | 3182   | 2  | RCU83443   | RCU83443 Rickettsia  |
| C 15       | 20    | 1.4   | 3182   | 2  | RCU83448   | RCU83448 Rickettsia  |
| C 16       | 20    | 1.4   | 3182   | 2  | RCU83453   | RCU83453 Rickettsia  |
| C 17       | 20    | 1.4   | 3182   | 2  | RUJ83442   | RUJ83442 Rickettsia  |
| C 18       | 20    | 1.4   | 3182   | 2  | RPB83449   | RPB83449 Rickettsia  |
| C 19       | 20    | 1.4   | 3182   | 2  | RSU83452   | RSU83452 Rickettsia  |
| C 20       | 20    | 1.4   | 3182   | 2  | RSU83454   | RSU83454 Rickettsia  |
| C 21       | 20    | 1.4   | 3185   | 2  | ARU83437   | ARU83437 Astrakhan r |
| C 22       | 20    | 1.4   | 3185   | 2  | TTU83441   | TTU83441 Israeli tic |
| C 23       | 20    | 1.4   | 3185   | 2  | RAU83446   | RAU83446 Rickettsia  |
| C 24       | 20    | 1.4   | 3188   | 2  | RMU83444   | RMU83444 Rickettsia  |

[illegible]

|             |         |             |       |              |
|-------------|---------|-------------|-------|--------------|
| Query Match | 100.0%; | Score 1442; | DB 5; | Length 1442; |
|-------------|---------|-------------|-------|--------------|

Query match 100.0%; score 1442;  
Best Local Similarity 100.0%; Pred. No. 0;

BEST LOCAL SIMILARITY 100.0%; FREQ: NO: 0;  
Matches 1442; Conservative 0; Mismatches

Qy 1 cccgggaactccatgtggccgtgatcctggtcgagcaggatatcgcatgatccagcggg 60

1 CCGGGAACTCCATGTGGCCGTGATCTCTGGTCCAGCAGGATATTCGGATGATCCAGCGGG 60

Qy 61 ccgcacagcgtgtgcggtaatggataaaggcctggttgttagaaacgctgacccaac 120

```
Db 1141 ATACACCGTGGCGGATGCTGAACAAACCTGTAGTTAGTAGGAATAACTAACCGG 1200
QY 1201 tgaacattaccggatgtagatcggggtaagtgtgaagtccaacaatcgctatttttaa 1260
Db 1201 TGAACATTACCGGATGTAGATCGGGTAATGTGAAGTTCAACAAATCGCTATTTTAA 1260
QY 1261 cagctaaagcagtgatcatatggggccagatacacaccatacaattatggtttactacc 1320
Db 1261 CAGCTAAAGCAGTGTCATATGGGGCCAGATACACCACCAATATTTGGTTTACTCTC 1320
QY 1321 ttacagcagtgacgcggcgcacaaagttgtcacaaatgcgcgagcaaccagactatt 1380
Db 1321 TTCACGGAGTGACCGGCGCACAAAGATTGTCAAAATGGCGGAGCAACCCAGGCTATT 1380
QY 1381 gccgaattaatcaaaatggcgcatcaacgcagaccactcaatgcaattcattggat 1440
Db 1381 GCCGAATTAAATCAAAATGGCGGCATCAACGCCAGACCACTCAATGCAATTCTTGGAT 1440
QY 1441 cc 1442
Db 1441 CC 1442

RESULT 2
LOCUS A72152 1442 bp DNA PAT 11-MAY-1999
DEFINITION Sequence 1 from Patent WO9801568.
ACCESSION A72152
VERSION A72152.1 GI:4808107
KEYWORDS
SOURCE Klebsiella oxytoca.
ORGANISM Klebsiella oxytoca
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Klebsiella.
REFERENCE 1 (bases 1 to 1442)
AUTHORS Brieden,W., Naughton,A., Robins,K., Shaw,N., Tinschert,A. and
Zimmermann,T.
TITLE METHOD OF PREPARING (S) - OR (R) -3,3,3-TRIFLUORO-2-HYDROXY-2-
METHYLPROPIONIC ACID
JOURNAL Patent: WO 9801568-A 15-JAN-1998;
LONZA AG (CH)
FEATURES
source
1..1442
Location/Qualifiers
/organism="Klebsiella oxytoca"
/strain="PR51"
/isolate="PR51"
/db_xref="taxon:571"
197..1183
/codon_start=1
/transl_table=11
/product="AMIDASE"
/protein_id="CAB42722.1"
/db_xref="GI:4808108"
/translation="MKWLEESIMAKRGVGRKPVTHLTEMQKEFHYTIGPVSTPV
LTIEPGRIVDTRDAFEGAINSEQDIPSOILKMPFLNPQNGIMVNGAEKGDVLAVY
TESLPRGVDPGICAMIPHEGLTGTDLTAMLNDPLPEKVRMIKIDSEKQVWSKRHT
LPYKPHIGTSLVSPEIDISINSLTPDNHGGNDVPDIPGSIITYLPVPAPGGRLEIGDA
HACODGEICCTAVEFASITTKVDLIKMWOLSWPMENAINISIGSARLEDA TRI
AYRLIYHLVDFEQQWDAYMLLSQCKVRLGNVDPKYTVGAMLNKNLLY"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN

Query Match 100.0%; Score 1442; DB 5; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgggaactccatggtggcggtgactcctgtgcgagcagatattgcgatgatccagcg 60
Db 1 CCGGGAACTCCATGTGGCGGTGATCTGTGTCGACGAGATATTGGATGATCCAGCGG 60
QY 61 ccgacagcgtgtgcggttaatggataaaggcctggtttagaacaacgtgacccaacac 120
|||||
```

```
Db 61 CCGCACAGCGCTGTGGGTAATGGATAAAGGCGCTGTGTTAGAAACGCTGACCCAAAC 120
QY 121 agctctctgatgatcttttaatgctgcgtcatctgctctgctgctgctgctgctgctgct 180
Db 121 AGCTCTCTCATGATCTTTTAATGCGTCTGTCATCTGCTCTGTAACCTAAACGCTATAAAT 180
QY 181 acgtggagaataacatatgaaatggttgagaataatccattatggccaaaacgoggtgttg 240
Db 181 AGCTGGAGAATAACATATGAAATGTTTGAAGAAATCCATTATGGCCAAACCGCGTGTGG 240
QY 241 tcccgggcgtaaacgggtaacgatacctgacggaagaataatgcaaaaagagtttcatta 300
Db 241 TCCCGGGCGTAAACCGGTAAACGATCACCTGACGGAGAAATGCAAAAAGAGTTTCATTA 300
QY 301 caccattggcccttattccacaccgctcctgaccatcgaaacccggtgacgcgattattgt 360
Db 301 CACCATTGGCCCTTATTCCACACCCGCTCTGACCATCGAAACCCGGTGACGGATATTGT 360
QY 361 cgacactcgagatgcttttgaagggtgctatcaattcggaacaggatatattccagagcagtt 420
Db 361 CGACACTCGAGATGCTTTTGAAGTGCTATCAATTGGGAACAGGATATTCCGAGCAGTT 420
QY 421 gctaaaaatgccttcttctcaaccccaaaaacggacgcgatacgtgcaatggcgcgagaa 480
Db 421 GCTAAAAATGCCCTTTCTCAACCCACAAAACGACGACGATCATGGTCAATGGCGCGAGAA 480
QY 481 aggtgatgtcgtcgtctctatatgaatccatgttgcctgcgcgctggtgatccctacgg 540
Db 481 AGGTGATGTGCTCGCTGCTCTATATCGAATCCATGTTGCCCGCGCGCTTGTATCCCTACGG 540
QY 541 catctgcgcctgattcccgcatcttggcggaactgacccgggacgcgacctgacgcgcaatgct 600
Db 541 CATCTGCGCCATGATTCGCCATTTTGGCGACTGACCGGACCGACCTGAGGCCATGCT 600
QY 601 caatgatcgcgtgccagaaaaagggtgcgcgtgattaaactcgacagtgaaaaggtctactg 660
Db 601 CAATGATCGCTGCCAGAAAAGGTGCGCATGATTAATAACTCGACAGTGAAGAAGTCTACTG 660
QY 661 gggcaaacgcctacgcttccctataaaccccatattggcaccttgagcgtatcgccaga 720
Db 661 GAGCAACGCCATACGCTTCCCTATNAACCCCATATTGGCACCTTGAGCCTATCGCCAGA 720
QY 721 aattgactcaatcaattcactgacgcagacaataacgcggggaatatggatgtgcgga 780
Db 721 AATTGACTCAATCAATTCACTGACGCGACACAATAACGCGGGAATATGGATGTGCGGA 780
QY 781 tataggaccaggagattacactatctgcgcgtacgtgcgcctggagcgcgcgtgttat 840
Db 781 TATAGGACCAGGAGATTACCTATCTGCCGTACCTGCCCTTGGAGCGCGCTGTTTAT 840
QY 841 tggatgcccctgctgtcagggtgagtgatgagatttgcggagccgcagtagagtttc 900
Db 841 TGGTGTGCCCATGCTGTGTCAGGGTGATGGTGAGATTTGCGGACCGCAGTAGAGTTTC 900
QY 901 ctcaatcacaccatcaaatgcgatttgatcaagaactggcagcttctctgcccacaa 960
Db 901 CTCATCACCAACATCAAAAGTCGATTGTATCAAGAACTGGCAGCTTCTCTGGCCACGAAT 960
QY 961 gagaaatgcgaaaaatattatgattgtgcagtcgcaatgcgcgtggagatgcgcgcg 1020
Db 961 GGAGAAATGCCGAAATATTATGAGTATTGCGAGTGACCGTCCCGTGGAGATGCGAGCG 1020
QY 1021 aattgcataatcggaacttaattactggtgtagagactttggttcgaacaatggga 1080
Db 1021 AATTGCATATCGCGACTTAATTTACTGGCTGGTAGAAGACTTTGGCTTCGAAACAATGGGA 1080
QY 1081 tgcctacatgcttctgagtcgaatgcggcaaatgcggcgtgggcaacatgctgcacccaa 1140
Db 1081 TGCTTACATGCTCTGAGTCAATCGCGCAAGTGGCGGTGGGCAACATGGTGCACCCCAA 1140
QY 1141 atacacgctgtgcgcgatctgaacaaaaacacctgttagtttagtaagaactaacaccg 1200
Db 1141 ATACACCGTGGCGGATGCTGAACAAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
```

```
Qy 1201 tgaacattaccggatgtagatcggggtaagtgttaagtccaacaatcgctatttttaa 1260
|||||
Db 1201 TGAACATTACCGGATGTAGATCGGGTAAATGTAAAGTTCACAAATCGCTATTTTAA 1260

Qy 1261 cagctaaagcgggtgcataatggtggccagatcacaccatcaatattggtttacttactcc 1320
|||||
Db 1261 CAGCTAAAGCAGGTGCATATGCGGCCAGATACACCCATCAATATTGTTACTTTACTTCC 1320

Qy 1321 ttcagcgagtgacgagcgcagagtggtcacaatggcgagcagcaaccagcgtatt 1380
|||||
Db 1321 TTCAGCGAGTGCAGCGGCCAGAGAGTGTCACAATGCGCGGAGCAACCCAGCGCTATT 1380

Qy 1381 gccgaataataatgagcggatcgaacgagcagaccactcaatgcaattcatttgat 1440
|||||
Db 1381 GCCGAATAATCAAAATGCGGCATCAACGGCAGACCACTCAATGCAATTCATTGGAT 1440

Qy 1441 cc 1442
||
Db 1441 CC 1442

RESULT 3
AC057459/c
LOCUS
DEFINITION Giardia intestinalis clone M2849 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
AC057459
VERSION AC057459.1 GI:7594157
KEYWORDS HTG; HTGS_PHASEO.
SOURCE
ORGANISM Giardia intestinalis.
REFERENCE
1 (bases 1 to 1100)
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Giardia: a model for ancient eukaryotic genome analysis
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1100)
AUTHORS Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2000) Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Marine Biological Laboratory, 7
MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT
* NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 1100: contig of 1100 bp in length.
FEATURES
source
Location/Qualifiers
1..1100
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clones="M2849"
BASE COUNT 268 a 299 c 222 g 311 t
ORIGIN

Query Match 1.4%; Score 20; DB 77; Length 1100;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 ccaacgcggtgtgtggtgcc 244
|||||
```

```
Db 229 CCAAACGCGGTGTTGGTGCC 210

RESULT 4
RMU83439/c
LOCUS
DEFINITION Rickettsia mongolotimonae rOmpA (ompA) gene, partial cds.
AC057439
VERSION U83439.1 GI:1778884
KEYWORDS
SOURCE Rickettsia mongolotimonae.
ORGANISM Rickettsia mongolotimonae
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiaceae; Rickettsia.
REFERENCE
1 (bases 1 to 3167)
AUTHORS Fournier,P.E., Roux,V. and Raoult,D.
TITLE Phylogenetic analysis of spotted fever group rickettsiae by study
of the outer surface protein rOmpA
JOURNAL Int. J. Syst. Bacteriol. 48 pt 3, 839-849 (1998)
MEDLINE 98404549
REFERENCE
2 (bases 1 to 3167)
AUTHORS Raoult,D., Fournier,P.-E. and Roux,V.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES
source
Location/Qualifiers
1..3167
/organism="Rickettsia mongolotimonae"
/strain="HA-91"
/db_xref="taxon:45261"
1..3167
/gene="ompA"
<1..>3167
/gene="ompA"
/codon_start=3
/transl_table=11
/product="rOmpA"
/db_xref="GI:1778885"
/db_xref="GI:1778885"
/translation="ITLHAGGSLDANNIDFGARSTLEFNGGGGIPYFKAIANGNNA
ILNVNKKLTAYHLTIGTVAEINIGAGNIFADASAGDVTILNADIDHFRALDSTLL
SNLTGCVNNILLAADLVAPGDEGTVFNGVNGNLNIGSVAGTARNIGGGGKNFN
ALLIYNAVITDDVLEIGNVLINNADFTSSTAFNAGAIQIONATYITIDANNGLN
IPAGNTOFAHADAOQLIQNSSGNDRTITLGANIDPDNDDEGIVILNSVTAGKLTIAQ
GKTGGAHKLQITVFKGAGDCSTAGTTFNTINIVLDTGOELGATVANYVLFKLDVQ
LTQGTIGGFLDFNAKGVTVLNNVNVAGTVQNTGNTNGTLVLGASNLNRVNGIA
MLKVGAGNVTIAKGNVKIGEIOGTNTLTLPAHNLGTSINKTGOALKLNFMSG
SVSGVGTAAANSVDITTAGATSFASVNAKGTAKLGTTSPAHNTTNGAVTLAKGS
ITSFARNVTATSEVANSATINFGNSLAFNSNITGSGTTLTLGANQVITYTGTGTDPL
TLNITFDGAAGSGNLIKSGSTLDLSGVSTLALVVTATNFMNNSPDPTKYTVISAE
TVGGLKPTKENVKITINDNRFVDFTFDASTLTLFAKIDAADVIDKDFAPGGLANI
PNAANIKKSLKLMEDAPNSDARFNNFGLMTPLOEADATTHLMODVVKPSTIAAV
NNOVASNTSSNITALNARMKVQAGNKGPVSSGDEMDAKFGAWISPFVGNATQKWC
NSLSGYKSDTGTGTGTFGDFGVDDLVGLAYTRADDIDKLKNNKTGDKKNVSNISYL
AGLISYPIENLTFVEATASDNKRSRRVIAVITLTETVGTGTANGKRYKSESTGQLM
AGTYMMPENINLTPLVGURYSYTIKDKGYKETGTTTQNTLVKGNKNTFDDLLGAKVS
SNINVEIVLTPELYAMVDYAFKNKYSAIDARLQGMTAPLPTNSFKQSKTSFDVGVG
TAKYKMWEGYINDYDNTIGSKYFAQQ"
BASE COUNT 1062 a 513 c 666 g 926 t
ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3167;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 aggcattatgcgcaataat 1392
|||||
Db 1603 AGGCTATTGCCGAATAAT 1584

RESULT 5
AF018076/c
```



```
LOCUS AF018076 3176 bp DNA BCT 08-DEC-1998
DEFINITION Rickettsia honei rOmpA (ompA) gene, partial cds.
ACCESSION AF018076
VERSION AF018076.1 GI:3982570
SOURCE Rickettsia honei.
ORGANISM Rickettsia honei.
REFERENCE 1 (bases 1 to 3176)
AUTHORS Stenos,J., Roux,V., Walker,D. and Raoult,D.
TITLE Island spotted fever in Australia
JOURNAL Rickettsia honei sp. nov., the aetiological agent of Flinders
MEDLINE Int. J. Syst. Bacteriol. 48, 1399-1404 (1998)
REFERENCE 99045882
AUTHORS Roux,V., Thiers,B. and Raoult,D.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1997) Faculte de Medecine, Unite des Rickettsies,
CNRS UPRES-A 6020, 27, bd Jean Moulin, Marseille 13385, France
FEATURES
source
1. .3176
/organism="Rickettsia honei"
/strain="RB"
/db_xref="taxon:37816"
<1. .>3176
/gene="ompA"
<1. .>3176
/gene="ompA"
/codon_start=3
/transl_table=11
/product="rOmpA"
/protein_id="AAC83454.1"
/db_xref="GI:3982571"
/translation="ITLQAGSLAANNIDFGAGSTLDENGLGGNAIPYYFKGAIAN
GNNAILNVNKLTAHYHLTIGTVAEINIGAGNLEIDAGSDVTILNQADINFRALDST
LVLSNLTGVGVNILLAAADLVAPGAGETVFNNGVNLGNISNVAQTARNIGDGGN
KNTLSIYNAVITDDVNLEGIQNLVINDNADFTSTAFNAGTQIKDATYITDANNG
FLNIPAGNIOFAHADAOQLILQNSGNDRTITLGANIDPDNDDEGIVILNSVTAGKILT
IAGCKTFGGAAHKLQTIIVFKGAGCGCTAGTFTNTNIVLDITGOLELGATVAVVLFND
AVOLTQTNIGGFLDPAKNGVTVLANNVAVCTVONTGNTNLTGSKTGQALKLNF
GIAMKVGAGNVTIAGKGVKIGIEQGTNTLTPAHFNLTSINTGSKGQALKLNF
NGSVGVGTAAVNSVDITAGTAFSSVNAKGTATLGTTSTFANTFTNTGAVTLA
KGSITFVKNATSVFANSATINFGNSLAFNSNITGSGTTLALVATNFDMNISPDKYTVI
DTLTLNFTDGAAGSGNLIKSGSTLDLSCVSTLALVATNFDMNISPDKYTVI
SATAGLTKTPKENVKTIINNDNRFVDFCDASTLFLPAEDIAADVIDEAPGGPL
ANIPNAANTIKKSLMEDAPNGSDARQANNFGLMTPLOEADATTHLMQDVVKPSDTI
AAYNQVVASNISSNITARNMDKVOAGNKGVPSSGDEMDAKFGAWISPFVSNATQ
KMCNSISGYKSDTGGTIGDFGVSDLLALGLAYTRADTDIKLKNKTKGDKNKVESNI
YSVGLYSPYENLFEVIAISYSDNKRISRRVIATLTETVGYQTANGRYKSESFTG
QLMAGTYMMPENINLTPLAGLRYSTIKDKGYKETGTYQNLTVKGNYNFDCGLIGA
KVSNNINVEILVTPELYAMVDYAFKNKVSADARLOGMTAPLPTNSFKSKTSFDFVG
VGYTAKHMEYGINDTNIGSKYFAQQ"
BASE COUNT 1063 a 523 c 677 g 913 t
ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3176;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 aggcattgccgaataaatt 1392
|||||
Db 1612 AGGCTATTGCCGAATAAT 1593

RESULT 6
RSU83455/c
LOCUS RSU83455 3176 bp DNA BCT 10-SEP-1998
DEFINITION Rickettsia sibirica rOmpA (ompA) gene, partial cds.
ACCESSION U83455
VERSION U83455.1 GI:1778904
KEYWORDS
```

```
LOCUS AF018076 3176 bp DNA BCT 10-SEP-1998
DEFINITION Rickettsia sibirica
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE 1 (bases 1 to 3176)
AUTHORS Fournier,P.E., Roux,V. and Raoult,D.
TITLE Phylogenetic analysis of spotted fever group rickettsiae by study
of the outer surface protein rOmpA
JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
REFERENCE 98404549
AUTHORS Raoult,D., Fournier,P.-E. and Roux,V.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES
source
1. .3176
/organism="Rickettsia sibirica"
/strain="246"
/db_xref="taxon:35793"
1. .3176
/gene="ompA"
<1. .>3176
/codon_start=3
/transl_table=11
/product="rOmpA"
/protein_id="AAC35191.1"
/db_xref="GI:1778905"
/translation="ITLHAGSLDANNIDFGARSTLEFNGLDGGIPYFKGAIAANG
NNAILNVNKLTAHYHLTIGTVAEINIGAGNIFADISAGDVTILNQADINFRALDST
LVLSNLTGVGVNILLAAADLVAPGAGETVFNNGVNLGHIGSVNAGTARNIGDGGN
KFNALLIYNAVITDDVNLEGIQNLVINDNADFTSTAFNAGTQIQNATYITDANNG
FLNIPAGNIOFAHADAOQLILQNSGNDRTITLGANIDPDNDDEGIVILNSVTAGKILT
IAGCKTFGGAAHKLQTIIVFKGAGCGCTAGTFTNTNIVLDITGOLELGATVAVVLFND
AVOLTQTNIGGFLDPAKNGVTVLNNVAVCTVONTGNTNLTGSKTGQALKLNF
GIAMKVGAGNVTIAGKGVKIGIEQGTNTLTPAHFNLTSINTGSKGQALKLNF
NGSVGVGTAAVNSVDITAGTAFSSVNAKGTATLGTTSTFANTFTNTGAVTLA
KGSITFVKNATSVFANSATINFGNSLAFNSNITGSGTTLALVATNFDMNISPDKYTVI
DTLTLNFTDGAAGSGNLIKSGSTLDLSCVSTLALVATNFDMNISPDKYTVI
SAAEAGLTKTPKENVKTIINNDNRFVDFCDASTLFLPAEDIAADVIDEAPGGPL
ANIPNAANTIKKSLMEDAPNGSDARQANNFGLMTPLOEADATTHLMQDVVKPSDTI
AAYNQVVASNISSNITARNMDKVOAGNKGVPSSGDEMDAKFGAWISPFVSNATQ
KMCNSISGYKSDTGGTIGDFGVSDLLALGLAYTRADTDIKLKNKTKGDKNKVESNI
YSVGLYSPYENLFEVIAISYSDIRIGKSRRIATLTETVGYQTANGRYKSESFTG
QLMAGTYMMPENINLTPVGLRYSTIKDKGYKETGTYQNLTVKGNYNFDCGLIGA
KVSNNINVEILVTPELYAMVDYAFKNKVSADARLOGMTAPLPTNSFKSKTSFDFVG
VGYTAKHMEYGINDTNIGSKYFAQQ"
BASE COUNT 1056 a 517 c 671 g 932 t
ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3176;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 aggcattgccgaataaatt 1392
|||||
Db 1612 AGGCTATTGCCGAATAAT 1593

RESULT 7
TTU83456/c
LOCUS TTU83456 3176 bp DNA BCT 10-SEP-1998
DEFINITION Thai tick typhus rickettsia rOmpA (ompA) gene, partial cds.
ACCESSION U83456
VERSION U83456.1 GI:1778906
KEYWORDS
SOURCE Thai tick typhus rickettsia.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE 1 (bases 1 to 3176)
```

AUTHORS Fournier, P.-E., Roux, V. and Raoult, D.  
 TITLE Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA  
 JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)  
 MEDLINE 98404549  
 REFERENCE 2 (bases 1 to 3176)  
 AUTHORS Raoult, D., Fournier, P.-E. and Roux, V.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies, CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France

FEATURES  
 Location/Qualifiers

source  
 1..3176  
 /organism="thai tick typhus rickettsia"  
 /strain="TT-118"  
 /db\_xref="taxon:45259"  
 gene  
 1..3176  
 /gene="ompA"  
 CDS  
 <1..>3176  
 /codon\_start=3  
 /transl\_table=11  
 /product="rOmpA"  
 /protein\_id="AAC35192.1"  
 /db\_xref="GI:1778907"

/translation="ITLQAGGSLAANNIDFGAGSTLDFNGLPGGGNAIPYFKGAIAN  
 GNNAIINVTKLTAYHSTIGTVAEINIGAGNLLIEDASAGDVTILNAQDINFRALDSA  
 LVLSNLTVGVGNLLAADLVAPGADEGTWVNGVGNLGINSNVAGTARNIGDGGGN  
 KNTLSIYNAVITDDVNLEGIQNVLINADNADFTSTFAGTIOIKDRTYIDANNG  
 NLNIPAGNTOFAHADQAQLIQNSGNDRTITLGANIDPDNDDEGIVILNSVAGKKLT  
 IAGKRFEGAHKLQITLVFKAGNCGAGTFTNTNIVLDITGOLELGATASVVLFDN  
 AVOLTOGTIGFGLDFNAKNGVTNNVNVNAGTVQNTGGTNGGTLILVLAGSNLRVN  
 GIAMLKVGAGNVTIAKGNVKIGETQGTNTLTPAHFNLTGSIKNTGGQALKLNF  
 NGSVSGVVTAAANSVDITTAGATSFASVNAKGTATLGGTTSFANTTNTGAVTLA  
 KGSITFAKNTATSPVANSATINFGNSLAFNSNITGSGTTLTLAGANVTYTGTSFT  
 DTLTNTFDGAAGSGNLLIKSGTLDLSGVSTLALVYATNFDMNNSPDKYVI  
 SAETAGLTAPKENVKTIINDNRFVDTFADSTLTTLFAEDIAADVIDEDFAPGGL  
 ANIPNAANIKKSLMEDAPNGSDARQAFNFGMLTPLQEDATTHLMQDVVVKPSDTI  
 KAVNNOVASNTSSNTALNARMKVQAGKGVSSGDEDMDAKFGATSPFVSNAIQ  
 KWCNSISGKTGTIGDFGVSDLLALGLAYTRADTDIKLKNKTKGKKVESNI  
 YSVGLNYPYENLFVEAJASDNKIRSKRRVIATLTETGYOTFANGKYKSEYTG  
 OLMAGVTYMPENILTPLAGLRYSTIKDKYKGTGTQNTLVKGNKYNPDGLLGA  
 KYSSNINVEILVPELYAMVDYAFANKYSAIDARLOGMTAPLPTNSFKQSKUSFDVG  
 VGVTAHKHMEYGINDTNIGSKYFAQQ"

BASE COUNT 1063 a 523 c 676 g 914 t  
 ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3176;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 agcgtattgccgaataat 1392  
 |||||  
 Db 1612 AGCTATTGCCGAATAAT 1593

RESULT 8  
 RMU83447/c  
 LOCUS RMU83447 3179 bp DNA BCT 10-SEP-1998  
 DEFINITION Rickettsia montana rOmpA (ompA) gene, partial cds.  
 ACCESSION U83447  
 VERSION U83447.1 GI:1778890  
 KEYWORDS  
 SOURCE Rickettsia montanensis.  
 ORGANISM Rickettsia montanensis.

REFERENCE  
 AUTHORS Fournier, P.-E., Roux, V. and Raoult, D.  
 TITLE Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA  
 JOURNAL Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)  
 MEDLINE 98404549

REFERENCE  
 AUTHORS Raoult, D., Fournier, P.-E. and Roux, V.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies, CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France

FEATURES  
 source

1..3179  
 /organism="Rickettsia montanensis"  
 /strain="M5/6"  
 /db\_xref="taxon:33991"  
 gene  
 1..3179  
 /gene="ompA"  
 CDS  
 <1..>3179  
 /codon\_start=3  
 /transl\_table=11  
 /product="rOmpA"  
 /protein\_id="AAC35183.1"  
 /db\_xref="GI:1778891"

/translation="ITLQAGGSLAANNIDFGVSTLEFNGLDGGDIIHYHFKGAIA  
 NGNNATLVNVTKSLTAYHSTIGTVAEINIGADSFETIDASAGDVTILNAQDINFRACN  
 STLMLNLTGVGNLLAADLVAPGADESCVVFNGMGLNIGSNVAGTARNIGDGG  
 GKFNNLFYINVVKYDIDVNLGKIVFIGNDADFTSSACNAGTQINNAYIAIDAN  
 NGNLNVPAGNIQFVHAGAOVLVLONSSENDRTIIGANIDPDNDGDGIVILNSVAGKKL  
 TIAGGKTFGGAHKLQAIVEFKAGNCGAGTFTNATDVLDTISQELGATTANVLLN  
 DAVLQTDIGDGLDFNARNGVTNNVNVNAGTVQNTGGTNGGTLILVLAGSNLSV  
 NGIAMLKVGAGNVTIAKGDVKIGETQGTNTLTPANFNLTGSIKNTGGQALKLNF  
 TNGSVSGVVTAAANSVDITTAGATSFASVNAKGTATLGGTTSFADTPTWTGAVTL  
 AKGSTINFAKNTATSPVANSATINFGNSLAFNSNITGSGTTLTLAGANVTYTGTSFT  
 TDTLTNTFDGAAGSGNLLIKSGTLDLSGVSTLALVYATNFDMNNSPDKYVI  
 ISAEIAGLKPTSKENVKTIINDNRFVDTFADSTLTTLFAEDIAADVIDEDFAPGGL  
 LANIPNAANIKKSLMEDAPNGSDARQAFNFGMLTPLQEDATTHLMQDVVVKPSDT  
 IAAVNOVASNTSSNTALNARMDEVAQNKGVSSGDEDMDAKFGATSPFVSNAIQ  
 KWCNNISGVKSDTGTIGDFGLSDLLALGLAYTRADTDIKLKNKMGDKKVESN  
 IYSLVGLYSPYENLFVEAJASDNKIRSKRRVIATLPTVGTQANGKYKSEYTT  
 GQMGAGTYMPENILTPLAGLRYSTIKDKYKGTGTQNTLVKGNKYNPDGLLGA  
 ARVSSNINVEILVPELYAMVDYAFANKYSAIDARLOGMAAPLPTNSFKQSKTSFDI  
 GVGVTAKHMEYGINDTNIGSKYFAQQ"

BASE COUNT 1055 a 516 c 682 g 926 t  
 ORIGIN

Query Match 1.4%; Score 20; DB 2; Length 3179;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 agcgtattgccgaataat 1392  
 |||||  
 Db 1615 AGCTATTGCCGAATAAT 1596

RESULT 9  
 AF120018S2/c

LOCUS AF120018S2 3182 bp DNA BCT 16-FEB-2000  
 DEFINITION Rickettsia sp. Dns28 OmpA (ompA) gene, partial cds.  
 ACCESSION AF120019  
 VERSION AF120019.1 GI:6979178  
 KEYWORDS  
 SEGMENT 2 of 2  
 SOURCE Rickettsia sp. Dns28.  
 ORGANISM Rickettsia sp. Dns28.

REFERENCE  
 AUTHORS Rydkina, E., Roux, V., Fetisova, N., Rudakov, N., Gafarova, M., Tarasevich, I. and Raoult, D.  
 TITLE Detection of New Spotted Fever Group Rickettsiae in Ticks Collected in the Former USSR  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3182)  
 AUTHORS Rydkina, E., Roux, V., Raoult, D. and Tarasevich, I.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JAN-1999) Unite Des Rickettsies, Faculte De Medecine,

```
FEATURES
  source
    27 Boulevard Jean Moulin, Marseille 13385, France
    Location/Qualifiers
      1. .3182
      /organism="Rickettsia sp. Dns28"
      /strain="Dns28"
      /db_xref="taxon:114649"
      /note="isolated in the former USSR"
      <1. .>3182
      /gene="ompA"
      /gene="ompA"
      /note="outer membrane protein"
      /codon_start=3
      /transl_table=11
      /product="OmpA"
      /protein_id="AAF34347.1"
      /db_xref="GI:6979180"
      /translation="ITLQAGGSLAANNIDFGVSTLEFNGPLDGGNTIPYFKGAIA
      NGNAILNVNTKSLTAYHSTIGTVAEINIGAGNLPALDASAGDVTILNAQDINFGAPD
      STLALSNTLGVGVKNILLAADLVAPGANAGDVVDFGGVNGNLNIGSNVAGTARNIGDGG
      GNKFTLLIYNAVITDDVNLEGIQNVLLNNADFTSSAFNAGAIQINDATYITIDAN
      NGNLVPAGNIQFAHADAKLILQNSGNDRTITLGANIDPDNDYEGIVTLNSVTAGKK
      LTIAGKTLGGAHKLQAIIVFKAGDFSAAGTFTNTVVDLTGQLELGATTAHVLL
      NDVOLTGTGNGGFLDFNAKGVTVLNNNVAGAVONTGTTNSGTLVLGASNLNS
      VNGIAMLKVGAGNVTIAKGDYKIGEIQGTGNTILPANFLGINSIKTSGOALKLN
      FTNGSVGVGTAAANSVDITAGTASFSSVNAKGTATLGGTTFADTFTNTGAVT
      LAKGSITFAKNTATSVANSATINFGSLAFNSITGSGTLLIGANQVYITGTGS
      FDMULTLNTFDGAVKSGNLIKSGNTLDLSGVSTLALVATNFIDNINISPDTKYT
      VISAETAGLKPTEENVKITINDNREFVDFTFASLTILFAEDATDVIDDFEPGG
      PLANTPNAANKIKSLELMEDAPNGSDARQAFNNFGLMTPLQPADATTHLMQDVVKPSD
      TIAANNVNASNISNTALNARKDKVOAANKGPVSSGDEMADAKFGWISPFVNGA
      TQKMCNNSGYKSDTGTGTFDGLVSDLLVLGLAYTRADDTIKLKNKTGDKNVES
      NIYSLGLYNVPYENLFEAIAFYSDNKRISRRVIATLETGVGTANGYKXSEY
      TQLMAGTYMMPENILTPLAGLRYSTIKDKYKETTQNTLVKGNKNTFQGLL
      GAKVSNINVEILVTPELYAMVDYAFKNKVPDAIDARLQGMTAPLPTNSFKQSKTSFD
      VGVGTAKHKMMEYGINVDNIGSKYFAQQ"
      BASE COUNT      1064 a      535 c      670 g      913 t
      ORIGIN

  Query Match      1.4%; Score 20; DB 2; Length 3182;
  Best Local Similarity 100.0%; Pred. No. 6.6;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1373 aggcatttcgcaataaataat 1392
      |||||||
  Db 1618 AGGCTATTGCCGAATAAT 1599

  RESULT 10
  AF120020S1/c
  LOCUS      AF120020S1 3182 bp DNA BCT 16-FEB-2000
  DEFINITION Rickettsia sp. Dns14 OmpA (ompA) gene, partial cds.
  ACCESSION  AF120020
  VERSION     AF120020.1 GI:6979182
  KEYWORDS
  SEGMENT
  SOURCE
  ORGANISM
    Rickettsia sp. Dns14.
    Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
    Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
    1 (bases 1 to 3182)
    Rydkina,E., Roux,V., Fetisova,N., Rudakov,N., Gafarova,M.,
    Tarasevich,I. and Raoult,D.
    Detection of New Spotted Fever Group Rickettsiae in Ticks Collected
    in the Former USSR
  JOURNAL
  REFERENCE
  2 (bases 1 to 3182)
  Rydkina,E., Roux,V., Raoult,D. and Tarasevich,I.
  AUTHORS
  TITLE
  Submitted (14-JAN-1999) Unite des Rickettsies, Faculte de Medecine,
  27 Boulevard Jean Moulin, Marseille 13385, France
  JOURNAL
  FEATURES
    Location/Qualifiers
      1. .3182
      /organism="Rickettsia sp. Dns14"
      /strain="Dns14"
      /db_xref="taxon:114648"
      /note="isolated in the former USSR"
      <1. .>3182
      /gene="ompA"
      /gene="ompA"
      /note="outer membrane protein"
      /codon_start=3
      /transl_table=11
      /product="OmpA"
      /protein_id="AAF34348.1"
      /db_xref="GI:6979184"
      /translation="ITLQAGGSLAANNIDFGVSTLEFNGPLDGGNTIPYFKGAIA
      NGNAILNVNTKSLTAYHSTIGTVAEINIGAGNLPALDASAGDVTILNAQDINFGAPD
      STLALSNTLGVGVKNILLAADLVAPGANAGDVVDFGGVNGNLNIGSNVAGTARNIGDGG
      GNKFTLLIYNAVITDDVNLEGIQNVLLNNADFTSSAFNAGAIQINDATYITIDAN
      NGNLVPAGNIQFAHADAKLILQNSGNDRTITLGANIDPDNDYEGIVTLNSVTAGKK
      LTIAGKTLGGAHKLQAIIVFKAGDFSAAGTFTNTVVDLTGQLELGATTAHVLL
      NDVOLTGTGNGGFLDFNAKGVTVLNNNVAGAVONTGTTNSGTLVLGASNLNS
      VNGIAMLKVGAGNVTIAKGDYKIGEIQGTGNTILPANFLGINSIKTSGOALKLN
      FTNGSVGVGTAAANSVDITAGTASFSSVNAKGTATLGGTTFADTFTNTGAVT
      LAKGSITFAKNTATSVANSATINFGSLAFNSITGSGTLLIGANQVYITGTGS
      FDMULTLNTFDGAVKSGNLIKSGNTLDLSGVSTLALVATNFIDNINISPDTKYT
      VISAETAGLKPTEENVKITINDNREFVDFTFASLTILFAEDATDVIDDFEPGG
      PLANTPNAANKIKSLELMEDAPNGSDARQAFNNFGLMTPLQPADATTHLMQDVVKPSD
      TIAANNVNASNISNTALNARKDKVOAANKGPVSSGDEMADAKFGWISPFVNGA
      TQKMCNNSGYKSDTGTGTFDGLVSDLLVLGLAYTRADDTIKLKNKTGDKNVES
      NIYSLGLYNVPYENLFEAIAFYSDNKRISRRVIATLETGVGTANGYKXSEY
      TQLMAGTYMMPENILTPLAGLRYSTIKDKYKETTQNTLVKGNKNTFQGLL
      GAKVSNINVEILVTPELYAMVDYAFKNKVPDAIDARLQGMTAPLPTNSFKQSKTSFD
      VGVGTAKHKMMEYGINVDNIGSKYFAQQ"
      BASE COUNT      1063 a      533 c      673 g      913 t
      ORIGIN

  Query Match      1.4%; Score 20; DB 2; Length 3182;
  Best Local Similarity 100.0%; Pred. No. 6.6;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1373 aggcatttcgcaataaataat 1392
      |||||||
  Db 1618 AGGCTATTGCCGAATAAT 1599

  RESULT 11
  AF120022S2/c
  LOCUS      AF120022S2 3182 bp DNA BCT 16-FEB-2000
  DEFINITION Rickettsia sp. RpA4 OmpA (ompA) gene, partial cds.
  ACCESSION  AF120023
  VERSION     AF120023.1 GI:6979188
  KEYWORDS
  SEGMENT
  SOURCE
  ORGANISM
    Rickettsia sp. RpA4.
    Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
    Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
    1 (bases 1 to 3182)
    Rydkina,E., Roux,V., Fetisova,N., Rudakov,N., Gafarova,M.,
    Tarasevich,I. and Raoult,D.
    Detection of New Spotted Fever Group Rickettsiae in Ticks Collected
    in the Former USSR
  JOURNAL
  REFERENCE
  2 (bases 1 to 3182)
    Rydkina,E., Roux,V., Raoult,D. and Tarasevich,I.
  AUTHORS
  TITLE
  Submitted (14-JAN-1999) Unite des Rickettsies, Faculte de Medecine,
  27 boulevard Jean Moulin, Marseille 13385, France
  JOURNAL
  FEATURES
    Location/Qualifiers
      1. .3182
      /organism="Rickettsia sp. RpA4"
      /strain="Dns14"
      /db_xref="taxon:114648"
      /note="isolated in the former USSR"
      <1. .>3182
      /gene="ompA"
      /gene="ompA"
      /note="outer membrane protein"
      /codon_start=3
      /transl_table=11
      /product="OmpA"
      /protein_id="AAF34348.1"
      /db_xref="GI:6979184"
      /translation="ITLQAGGSLAANNIDFGVSTLEFNGPLDGGNTIPYFKGAIA
      NGNAILNVNTKSLTAYHSTIGTVAEINIGAGNLPALDASAGDVTILNAQDINFGAPD
      STLALSNTLGVGVKNILLAADLVAPGANAGDVVDFGGVNGNLNIGSNVAGTARNIGDGG
      GNKFTLLIYNAVITDDVNLEGIQNVLLNNADFTSSAFNAGAIQINDATYITIDAN
      NGNLVPAGNIQFAHADAKLILQNSGNDRTITLGANIDPDNDYEGIVTLNSVTAGKK
      LTIAGKTLGGAHKLQAIIVFKAGDFSAAGTFTNTVVDLTGQLELGATTAHVLL
      NDVOLTGTGNGGFLDFNAKGVTVLNNNVAGAVONTGTTNSGTLVLGASNLNS
      VNGIAMLKVGAGNVTIAKGDYKIGEIQGTGNTILPANFLGINSIKTSGOALKLN
      FTNGSVGVGTAAANSVDITAGTASFSSVNAKGTATLGGTTFADTFTNTGAVT
      LAKGSITFAKNTATSVANSATINFGSLAFNSITGSGTLLIGANQVYITGTGS
      FDMULTLNTFDGAVKSGNLIKSGNTLDLSGVSTLALVATNFIDNINISPDTKYT
      VISAETAGLKPTEENVKITINDNREFVDFTFASLTILFAEDATDVIDDFEPGG
      PLANTPNAANKIKSLELMEDAPNGSDARQAFNNFGLMTPLQPADATTHLMQDVVKPSD
      TIAANNVNASNISNTALNARKDKVOAANKGPVSSGDEMADAKFGWISPFVNGA
      TQKMCNNSGYKSDTGTGTFDGLVSDLLVLGLAYTRADDTIKLKNKTGDKNVES
      NIYSLGLYNVPYENLFEAIAFYSDNKRISRRVIATLETGVGTANGYKXSEY
      TQLMAGTYMMPENILTPLAGLRYSTIKDKYKETTQNTLVKGNKNTFQGLL
      GAKVSNINVEILVTPELYAMVDYAFKNKVPDAIDARLQGMTAPLPTNSFKQSKTSFD
      VGVGTAKHKMMEYGINVDNIGSKYFAQQ"
      BASE COUNT      1063 a      533 c      673 g      913 t
      ORIGIN
```



```
/product="OmpA"
/db_protein_id="AAC35176.1"
/db_xref="GI:1778875"
/translation="IIQAAGSLDANNIDFGARSTLEFNGPLDGGGNAIPYVFKGAIA
NGNAILNVNTKLLTAYHLTIGTVAEINIGAGNLFDAIDASAGDVTILNAQDIHFRALD
SALVLSNLTCGVGNVLLAALVAPGVDEGTVPFDGGVNGNLGNSVAGAAARNIGDVG
GNKFNTHLLIYNAVITDDVNLEGIONVLNNADFTSSTAFNAGTIQINDATYTTIDAN
KGNLNPAGNIKFAHAQLOAIVFKGEGDFGTAGTTNTNIVLDITGQLELGATTANVVL
KDAVOLTOGTNGLGGFLDFNAKNGVTTLNNVNAVGTVKNTGTTNGTTLVLGASNLNR
VNGIAMLKVGAGNVTIAGGNVKGIEIOGTGNTLTLPAHFKLGTSGINTGGOALKLN
FMNGSGVGVGTAAVNSVDITTAGATSPASSNAKGTATLGTTSTFAHTFTNTGAVT
LAKGSIISFAKNVTATSFVANSATINFGNSLAFNSITGTTTLTGANQVYTTGTS
FTDITLTLTLDGAAKSGNLIKSGSTLDLSGVSNLALVVTATNFMNNSPDTKYT
VISATAGLTKPTKENVKITINNDRFVDFDASTLTLFAEDIAAGVIDEDFAPGG
PLANTPNAANIKKSLELMEDAPNGSDAQAFNFGMLTPOEDATTHLMQDVYKPSD
TIAAVNNQVVASNISNITARNMDKVOAGNKGVPSSGDEMDAKFCAWISPFVGN
TOKMNSISGVKSDTGTGTFGVSDDLVLGLAYTRADTIDIKLNKKTGDKNKVES
NIYSLYGIVPYENLVEAIASDNKIRSKRRVIATTLTETVGTQFANGYKSES
TGQLMAGTYIMSENINLTPLAGLRYSTIKDKSYKETGTYQNTLVKGNYNTEFGLL
GAKVSSNNINVEIIVLPELYAMVDYAFKNKVSADARLQGMATPLPTNSFKQKTSFD
VGVGVTAKHKMXYGINYDNTNIGSKYFAQQ"
BASE COUNT      1065 a      510 c      673 g      934 t
ORIGIN

Query Match      1.4%; Score 20; DB 2; Length 3182;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 aggcattgcgaataat 1392
|||||
DB 1618 AGGCTATTGCCGAATAAT 1599

RESULT 14
RCU83443/c
LOCUS      Rickettsia conorii romPA (ompA) gene, partial cds.
DEFINITION
ACCESSION  U83443
VERSION     U83443.1 GI:1778876
KEYWORDS   Rickettsia conorii.
SOURCE     Rickettsia conorii.
ORGANISM   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiidae; Rickettsia.
REFERENCE  1 (bases 1 to 3182)
AUTHORS    Fournier,P.E., Roux,V. and Raoult,D.
TITLE      Phylogenetic analysis of spotted fever group rickettsiae by study
            of the outer surface protein romPA
JOURNAL    Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
MEDLINE    98404549
REFERENCE  2 (bases 1 to 3182)
AUTHORS    Raoult,D., Fournier,P.-E. and Roux,V.
TITLE      Direct Submission
JOURNAL    Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
            CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES    Location/Qualifiers
            source
            1..3182
            /organism="Rickettsia conorii"
            /strain="M1"
            /db_xref="taxon:781"
            1..3182
            /gene="ompA"
            <1..>3182
            /gene="ompA"
            /codon_start=3
            /transl_table=11
            /product="OmpA"
            /protein_id="AAC35179.1"
            /db_xref="GI:1778877"
            /translation="VTLOAGSLDANNIDFGARSTLEFNGPLDGGGNAIPYVFKGAIA
            NGNAILNVNTKLLTAYHLTIGTVAEINIGAGNLFDAIDASAGDVTILNAQDIHFRALD
            SALVLSNLTCGVGNVLLAALVAPGVDEGTVPFDGGVNGNLGNSVAGAAARNIGDVG
            GNKFNTHLLIYNAVITDDVNLEGIONVLNNADFTSSTAFNAGTIQINDATYTTIDAN
            KGNLNPAGNIKFAHAQLOAIVFKGEGDFGTAGTTNTNIVLDITGQLELGATTANVVL
            KDAVOLTOGTNGLGGFLDFNAKNGVTTLNNVNAVGTVKNTGTTNGTTLVLGASNLNR
            VNGIAMLKVGAGNVTIAGGNVKGIEIOGTGNTLTLPAHFKLGTSGINTGGOALKLN
            FMNGSGVGVGTAAVNSVDITTAGATSPASSNAKGTATLGTTSTFAHTFTNTGAVT
            LAKGSIISFAKNVTATSFVANSATINFGNSLAFNSITGTTTLTGANQVYTTGTS
            FTDITLTLTLDGAAKSGNLIKSGSTLDLSGVSNLALVVTATNFMNNSPDTKYT
            VISATAGLTKPTKENVKITINNDRFVDFDASTLTLFAEDIAAGVIDEDFAPGG
            PLANTPNAANIKKSLELMEDAPNGSDAQAFNFGMLTPOEDATTHLMQDVYKPSD
            TIAAVNNQVVASNISNITARNMDKVOAGNKGVPSSGDEMDAKFCAWISPFVGN
            TOKMNSISGVKSDTGTGTFGVSDDLVLGLAYTRADTIDIKLNKKTGDKNKVES
            NIYSLYGIVPYENLVEAIASDNKIRSKRRVIATTLTETVGTQFANGYKSES
            TGQLMAGTYIMSENINLTPLAGLRYSTIKDKSYKETGTYQNTLVKGNYNTEFGLL
            GAKVSSNNINVEIIVLPELYAMVDYAFKNKVSADARLQGMATPLPTNSFKQKTSFD
            VGVGVTAKHKMXYGINYDNTNIGSKYFAQQ"
```

```
SALVLSNLTCGVGNVLLAALVAPGVDEGTVPFDGGVNGNLGNSVAGAAARNIGDVG
GNKFNTHLLIYNAVITDDVNLEGIONVLNNADFTSSTAFNAGTIQINDATYTTIDAN
KGNLNPAGNIKFAHAQLOAIVFKGEGDFGTAGTTNTNIVLDITGQLELGATTANVVL
KDAVOLTOGTNGLGGFLDFNAKNGVTTLNNVNAVGTVKNTGTTNGTTLVLGASNLNR
VNGIAMLKVGAGNVTIAGGNVKGIEIOGTGNTLTLPAHFKLGTSGINTGGOALKLN
FMNGSGVGVGTAAVNSVDITTAGATSPASSNAKGTATLGTTSTFAHTFTNTGAVT
LAKGSIISFAKNVTATSFVANSATINFGNSLAFNSITGTTTLTGANQVYTTGTS
FTDITLTLTLDGAAKSGNLIKSGSTLDLSGVSNLALVVTATNFMNNSPDTKYT
VISATAGLTKPTKENVKITINNDRFVDFDASTLTLFAEDIAAGVIDEDFAPGG
PLANTPNAANIKKSLELMEDAPNGSDAQAFNFGMLTPOEDATTHLMQDVYKPSD
TIAAVNNQVVASNISNITARNMDKVOAGNKGVPSSGDEMDAKFCAWISPFVGN
TOKMNSISGVKSDTGTGTFGVSDDLVLGLAYTRADTIDIKLNKKTGDKNKVES
NIYSLYGIVPYENLVEAIASDNKIRSKRRVIATTLTETVGTQFANGYKSES
TGQLMAGTYIMSENINLTPLAGLRYSTIKDKSYKETGTYQNTLVKGNYNTEFGLL
GAKVSSNNINVEIIVLPELYAMVDYAFKNKVSADARLQGMATPLPTNSFKQKTSFD
VGVGVTAKHKMXYGINYDNTNIGSKYFAQQ"
BASE COUNT      1064 a      511 c      674 g      933 t
ORIGIN

Query Match      1.4%; Score 20; DB 2; Length 3182;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 aggcattgcgaataat 1392
|||||
DB 1618 AGGCTATTGCCGAATAAT 1599

RESULT 15
RCU83448/c
LOCUS      Rickettsia conorii romPA (ompA) gene, partial cds.
DEFINITION
ACCESSION  U83448
VERSION     U83448.1 GI:1778878
KEYWORDS   Rickettsia conorii.
SOURCE     Rickettsia conorii.
ORGANISM   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiidae; Rickettsia.
REFERENCE  1 (bases 1 to 3182)
AUTHORS    Fournier,P.E., Roux,V. and Raoult,D.
TITLE      Phylogenetic analysis of spotted fever group rickettsiae by study
            of the outer surface protein romPA
JOURNAL    Int. J. Syst. Bacteriol. 48 Pt 3, 839-849 (1998)
MEDLINE    98404549
REFERENCE  2 (bases 1 to 3182)
AUTHORS    Raoult,D., Fournier,P.-E. and Roux,V.
TITLE      Direct Submission
JOURNAL    Submitted (31-DEC-1996) Faculte de Medecine, Unite des Rickettsies,
            CNRS EP J0054, 27 bd Jean Moulin, Cedex 5, Marseille, France
FEATURES    Location/Qualifiers
            source
            1..3182
            /organism="Rickettsia conorii"
            /strain="Moroccan"
            /db_xref="taxon:781"
            1..3182
            /gene="ompA"
            <1..>3182
            /gene="ompA"
            /codon_start=3
            /transl_table=11
            /product="OmpA"
            /protein_id="AAC35184.1"
            /db_xref="GI:1778879"
            /translation="VTLOAGSLDANNIDFGARSTLEFNGPLDGGGNAIPYVFKGAIA
            NGNAILNVNTKLLTAYHLTIGTVAEINIGAGNLFDAIDASAGDVTILNAQDIHFRALD
            SALVLSNLTCGVGNVLLAALVAPGVDEGTVPFDGGVNGNLGNSVAGAAARNIGDVG
            GNKFNTHLLIYNAVITDDVNLEGIONVLNNADFTSSTAFNAGTIQINDATYTTIDAN
            KGNLNPAGNIKFAHAQLOAIVFKGEGDFGTAGTTNTNIVLDITGQLELGATTANVVL
            KDAVOLTOGTNGLGGFLDFNAKNGVTTLNNVNAVGTVKNTGTTNGTTLVLGASNLNR
```



| Result<br>No. | Score | Query |        |   | DB     | ID                 | Description |
|---------------|-------|-------|--------|---|--------|--------------------|-------------|
|               |       | Match | Length | % |        |                    |             |
| 1             | 1442  | 100.0 | 1442   | 1 | N10449 | K. oxytoca R-speci |             |
| C 2           | 18    | 1.2   | 1263   | 1 | V82205 | Nad B gene encodin |             |
| C 3           | 18    | 1.2   | 1442   | 1 | V10449 | K. oxytoca R-speci |             |
| C 4           | 18    | 1.2   | 1724   | 1 | Q03361 | The nadA nucleotid |             |
| C 5           | 18    | 1.2   | 1724   | 1 | Q03363 | The nadA nucleotid |             |
| C 6           | 17    | 1.2   | 26     | 1 | Q49599 | lacs gene C-termin |             |
| C 7           | 17    | 1.2   | 26     | 1 | Q49600 | lacs gene C-termin |             |
| C 8           | 17    | 1.2   | 26     | 1 | Q51468 | Primer 2. Stable i |             |
| C 9           | 17    | 1.2   | 26     | 1 | Q51469 | Primer 3. Stable i |             |
| 10            | 17    | 1.2   | 1196   | 1 | Q20812 | Human haematopoiet |             |
| 11            | 17    | 1.2   | 1501   | 1 | V00522 | HIV-1 breakthrough |             |
| 12            | 17    | 1.2   | 1512   | 1 | V00521 | HIV-1 breakthrough |             |
| 13            | 17    | 1.2   | 1883   | 1 | X20225 | Enterococcus faeca |             |
| 14            | 17    | 1.2   | 2019   | 1 | X20306 | Borrelia burgdorfe |             |
| C 15          | 17    | 1.2   | 3800   | 1 | X28141 | CCoAOMT promoter,  |             |
| C 16          | 17    | 1.2   | 3963   | 1 | X20224 | Enterococcus faeca |             |
| 17            | 17    | 1.2   | 3963   | 1 | X20226 | Enterococcus faeca |             |
| C 18          | 17    | 1.2   | 4305   | 1 | X20275 | Borrelia burgdorfe |             |
| C 19          | 17    | 1.2   | 4312   | 1 | X20274 | Borrelia burgdorfe |             |
| C 20          | 17    | 1.2   | 5558   | 1 | X12948 | Enterococcus faeca |             |
| C 21          | 17    | 1.2   | 6444   | 1 | V74544 | Staphylococcus aur |             |
| 22            | 17    | 1.2   | 7227   | 1 | N60194 | Sequence encoding  |             |
| 23            | 17    | 1.2   | 9813   | 1 | X13059 | Enterococcus faeca |             |
| 24            | 16    | 1.1   | 99     | 1 | Q21267 | PND EE322-1. New a |             |
| 25            | 16    | 1.1   | 102    | 1 | Q21239 | PND EE244-2. New a |             |
| 26            | 16    | 1.1   | 102    | 1 | Q21236 | PND EE229-2. New a |             |
| 27            | 16    | 1.1   | 102    | 1 | Q21235 | PND EE229-1. New a |             |
| 28            | 16    | 1.1   | 102    | 1 | Q21237 | PND EE229-3. New a |             |
| 29            | 16    | 1.1   | 102    | 1 | Q21238 | PND EE244-1. New a |             |
| 30            | 16    | 1.1   | 102    | 1 | Q21240 | PND EE244-3. New a |             |
| 31            | 16    | 1.1   | 102    | 1 | Q21371 | PND EE1476-1. New  |             |
| 32            | 16    | 1.1   | 105    | 1 | Q21309 | PND EE505-2. New a |             |
| 33            | 16    | 1.1   | 105    | 1 | Q21310 | PND EE505-3. New a |             |

```
db 181 ACCTGAGAAATACATATGAATGGTTGGAGAAATCCATTATGCGCAAAACGGGTGTTGG 240
QY 241 tgcggggttaacgggttaacgcatcacctgacggaagaataagaaagatttcatta 300
Db 241 TGCggggGtTaaCGGtTaaCGcATcAcCTGAGCGAAAGAAATGCAAAAGAGATTTCATTA 300
QY 301 caccattggcccttattccacaacccgctcctgaccatcgaaaccccggtgaccggattattgt 360
Db 301 CACCATTGGCCCTTATTCCACACCGCTCTGACCATCGAATCCCGGTGACCGGATTATTGT 360
QY 361 cgacactgagatgcttttgaaggtgctatcaattcgaaacagagatattccagccaggt 420
Db 361 CGACACTCGAGATGCTTTGAAGGTGCTATCAATTCGAAACAGGATATTCGAGCCAGTT 420
QY 421 gctaaaaatgccccttctcaaccacacaaagcggaccatcatgctcaatggcggagaa 480
Db 421 GCTAAAAATGCCCTTCTCAACCCACAAAACGGACCGATCATGTGCTCAATGGCGGGAGAA 480
QY 481 aggtgatgctcgtcgtctctatatcgatccatgcttcccgcgcggttgatccctacgg 540
Db 481 AGGTGATGCTCGCTGCTATATCGAATCCATGTTGCCCGCGGGGTTGATCCCTACGG 540
QY 541 catctggcgaatgattcccgatcttggcgactgaccgggacccgaccctgacgcccattgct 600
Db 541 CATCTGGCGCATGATTCGGATTTTGGCGGACTGACCGGGACCGACCTGACGCGCCATGCT 600
QY 601 caatgatcccgctgcagaaaaagtgccgatgattaaactcgacagtgaaaggtctactg 660
Db 601 CAATGATCCGCTGCCAGAAAAGTGCATGATTAACTCGACAGTGAAGGTCTACTG 660
QY 661 gacgaacgcatcactgcttccctataaaccccatattggcacccttgagcgtatccacga 720
Db 661 GAGCAACGCGATACGCTTCCCTATAAACCCCATATTGGCACCTTGAGCGTATCGCCAGA 720
QY 721 aattgactcaatcaatcactgacgcagacaaatcacgcggggaatatggtgctccgga 780
Db 721 AATTGACTCAATCAATCACTGACGCGCAGACAAATCACGCGGGAATATGATGTGCCGA 780
QY 781 tatgagaccaggagattaccctatctgcgtacgtgcgctgacgctgagaccctcttat 840
Db 781 TATGAGACCAGGAGTATTACCTATCTGCGGTAGCTGCGCTGAGCGCGGCTCTTTAT 840
QY 841 tggtagtcccctgcttgcagggtgtagtgagattgcgggaccgagtagagttgc 900
Db 841 TGGTAGTCCCCTGCTTGTGAGGGTGTGTTGAGATTTGCGGGACCGCAGTAGAGTTGC 900
QY 901 ctcaatcaaccacatcaaatgctattgatcaagaactgagcagcttctctggccacgaat 960
Db 901 CTCAATCACCAACATCAAGTCGATTTCATCAAGNACTGGCAGCTTCTCTGGCCACCAAT 960
QY 961 ggagaatccgaaaaatattatgattggtgagtgacgtccgctgagggatgcgacg 1020
Db 961 GGAGAAATCCGAAAATATTATGATTTGGCAGTGCACGTCGCTGGAGGATGGACGG 1020
QY 1021 aattgcatatcgacttaatttactgctggttagaagacttggctcgaaacatggga 1080
Db 1021 AATTGCATATCGACITTAATTTACTGGCTGGTAGAAGACTTTGGCTTCGAACAATGGGA 1080
QY 1081 tgcctaatgcttctgagtcgaatcgcgcaaaagtgcgctgggcaacatgctgaccccaa 1140
Db 1081 TGCCTATATGCTTCTGAGTCAATGCGGCAAAAGTGCAGTGGCAACATGCTGACCCCAA 1140
QY 1141 atacaccgttggcgagatgctgaacaaaaaccgttagtttagtaggaataactaacccgg 1200
Db 1141 ATACACCGTTGGCGGATGCTGAACAAAACCTGTAGTTTAGTAGGAATAACTAACCGG 1200
QY 1201 tgaacattaccggatgtagatcggggtgaatgtgtaagtttcaaacatcgctatttttaa 1260
Db 1201 TGAACATTACCAGGATGATAGATCGGGTAATGTGTAAGTTCAAACAATCGCTATTTTAA 1260
QY 1261 cagctaaagcagtgcatatggggccagatacacccatcaattggtttacttactoc 1320
Db 1261 CAGCTAAAGCAGTGTATATGGGGCCAGATACACCCATCAATATTTGGTTTACTTTACTCC 1320
```

```
QY 1321 ttcagcgagtgacgcggcacaaagattgtcacaatggcggagcaacccaggctatt 1380
Db 1321 TTCAGCGAGTgACGCGGCGCAAGAGTTGTCAATGCGCGGAGCAACCCAGGCTATT 1380
QY 1381 gccgaattaatcaaatggcgcatcaacggcagaccactcaatcaattcatttggat 1440
Db 1381 GCCGAAATTAATCAAAATGGCGCATCAACGCGAGACCTCAATGCAATTCATTGGAT 1440
QY 1441 cc 1442
Db 1441 CC 1442
```

```
RESULT 2
N82205/c
ID N82205 standard; DNA; 1263 BP.
AC N82205;
DT 12-NOV-1990 (first entry)
DE Nad B gene encoding L aspartate 3' end.
KW Quinolinic acid synthase; L aspartate oxidase; nad A; nad B;
OS plant protection agent; ss.
PN Escherichia coli (K12C600).
PD DE3703255-A.
PF 18-AUG-1988.
PR 04-FEB-1987; 703255
PA 04-FEB-1987; DE-703255.
PI (RUTG) Rurgerswerke AG.
PI Laufer A, Gassen HG, Flachmann R, Hoke H, Holtmann W, Kunz N,
PI Stadelhofer J, Seifert J;
DR WPI; 88-235937/34.
PT Plasmids coded for quinolinic acid synthesis - have DNA sequences contg.
PT genetic information for synthesis of quinolinic synthase and L aspartate
PT oxidase.
PS Claim 7; Page 4-5; 24pp; German.
CC Between this 3' end sequence and the 5' end sequence of N80035 are
CC ca. 200 bp not represented in the specification.
CC The nad A sequence (N80035) encoding a polypeptide with the biological
CC activity of quinolinic acid synthase, together with nad B encoding a
CC polypeptide with the biological activity of L aspartate is inserted into
CC a plasmid. The quinolinic acid expressed is an intermediate for
CC pharmaceuticals and plant protection agents.
SQ Sequence 1263 BP; 307 A; 324 C; 344 G; 284 T; 4 Others;
```

```
Query Match 1-2%; Score 18; DB 1; Length 1263;
Best Local Similarity 100.0%; Fred. No. 7.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 459 tcattgtcaatggcgcg 476
Db 594 TCATGTCATATGGCGGG 577
```

```
RESULT 3
V10449/c
ID V10449 standard; DNA; 1442 BP.
AC V10449;
DT 16-JUN-1998 (first entry)
DE K. oxytoca R-specific amidohydrolase genomic DNA.
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform; ds.
OS Klebsiella oxytoca.
FH Key Location/Qualifiers
FT CDS
FT 197..1183
FT /*tag= a
FT /product=
FT /note= "R-specific amidohydrolase"
PN W09801568-A2.
PD 15-JAN-1998.
PE 10-JUL-1997; E03670.
PR 03-MAR-1997; CH-000500.
PR 10-JUL-1996; CH-001723.
```



PA (LONZ ) LONZA AG.  
 PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,  
 PI Zimmermann T;  
 WPI: 98-101063/09.  
 DR P-PSDB: W40263.  
 PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation  
 PT - by stereoselective hydrolysis of corresponding racemic amide using  
 PT microorganism or derived enzyme, used as drug intermediate  
 PS Claim 7a; Page 32-34; 68pp; German.  
 CC This DNA sequence encodes a R-specific amidohydrolase isolated from  
 CC Klebsiella oxytoca strain PRS1 which allows the microorganism to  
 CC utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole  
 CC nitrogen source. This amidohydrolase is used in a process for preparing  
 CC (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which  
 CC is cheaper than prior art optical resolution of the racemate using  
 CC dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.  
 SQ Sequence 1442 BP; 385 A; 350 C; 370 G; 337 T;

Query Match 1.2%; Score 18; DB 1; Length 1442;  
 Best Local Similarity 100.0%; Pred. No. 7.4;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 cgcacagcgtgtgcgg 78  
 Db 78 CGCACAGCGCTGCGG 61

RESULT 4  
 Q03361/C  
 ID Q03361 standard; DNA; 1724 BP.  
 AC Q03361;  
 DE 24-JUL-1990 (first entry)  
 DE The nadA nucleotide sequence used in effecting prodn. of quinolinic acid  
 KW Quinolinic acid production; nadA; nadB; pharmaceutical intermediate;  
 KW Plant protection agent intermediate; nadA; nadB; pharmaceutical intermediate;  
 KW L-aspartate oxidase; Escherichia coli W4546; quinolinic acid synthase;  
 PN DE3826040-A.  
 PD 08-FEB-1990.  
 PF 30-JUL-1988; 826040.  
 PR 30-JUL-1988; DE-826040.  
 PA (RUTG) Rutgerswerke Ag.  
 PI Laufer A, Gassen HG, Flachmann R, Kunz N;  
 WPI: 90-045655/07.  
 DR Fermentative prodn. of quinolinic acid -  
 PT using genetically transformed microorganisms contg. nadA and  
 PT nadB genes  
 PS Disclosure; p; German.  
 CC It codes for quinolinic acid synthase. Also claimed are nadB which codes  
 CC for L-aspartate oxidase and genetically transformed microorganisms contg.  
 CC plasmids bearing nadA and nadB in a copy ratio of 200:50 or 200:8.  
 CC Transformation of E. coli W4546 successively with pQAB 520 and pQAA 168  
 CC produced a transformant with a 200:50 copy ratio, giving quinolinic acid  
 CC in yields which were ca. 7.5-9.7 times those obtained using W4546. The  
 CC microorganisms give high yields in a reproducible manner. Quinolinic  
 CC acid is an intermediate for pharmaceuticals and plant protection agents.  
 SQ Sequence 1724 BP; 408 A; 432 C; 470 G; 414 T;

Query Match 1.2%; Score 18; DB 1; Length 1724;  
 Best Local Similarity 100.0%; Pred. No. 7.5;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 tcattgtcaatggcgcgg 476  
 Db 976 TCATGTCATGCGCGG 959

RESULT 5  
 Q03363/C  
 ID Q03363 standard; DNA; 1724 BP.  
 AC Q03363;  
 DT 24-JUL-1990 (first entry)

DE The nadA nucleotide sequence used in effecting prodn. of quinolinic acid  
 KW Quinolinic acid production; nadA; nadB; pharmaceutical intermediate;  
 KW Plant protection agent intermediate; nadA; nadB; pharmaceutical intermediate;  
 KW L-aspartate oxidase; Escherichia coli W4546; Escherichia coli NK6042;  
 KW pQAB510; pQAB520; pQAB203; pHT234; pQAB104; pLG339; Escherichia coli RFI;  
 KW pQAA12; pUC18; pQAA166; pQAA169.  
 PN DE3826041-A.  
 PD 08-FEB-1990.  
 PF 30-JUL-1988; 826041.  
 PR 30-JUL-1988; DE-826041.  
 PA (RUTG) Rutgerswerke Ag.  
 PI Laufer A, Gassen HG, Flachmann R, Kunz N, Seifert J;  
 WPI: 90-045656/07.  
 DR Fermentative prodn. of quinolinic acid -  
 PT using genetically transformed microorganisms contg. nadA and  
 PT nadB genes  
 PS Disclosure; p; German.  
 CC It codes for quinolinic acid synthase. Also claimed are nadB which codes  
 CC for L-aspartate oxidase and genetically transformed microorganisms contg.  
 CC plasmids bearing nadA and nadB DNA in a copy ratio of 50:200, 8:50 or  
 CC 8:200. Isolated chromosomal DNA was digested with HindII to obtain 6 and  
 CC 8 kb fragments, which were inserted into pBR322 and cloned in E. coli  
 CC NK6042 to obtain plasmid pQAB510. This was digested with HindIII and NruI  
 CC to obtain a 3.2 kb fragment, which was inserted in pBR322 to produce pQAB  
 CC 520. This was digested with SspI and NruI to obtain a 2.5 kb fragment,  
 CC which was inserted in pUC18 to obtain pQAB203. This was digested with  
 CC PstI and BamHI to give a 2.5 kb fragment, which was inserted into pHT234  
 CC to give the nadB-contg. plasmid pQAB104. The same chromosomal DNA was  
 CC digested with BamHI, inserted into pLG339 and cloned in E. coli RFI to  
 CC obtain pQAA12. This was digested with AluI, ligated with HincII-cleaved  
 CC pUC18 and cloned in E. coli 431 to give pQAA166. This was digested with  
 CC BamII and PstI to isolate a 1.4 kb insert, which was inserted into psti/  
 CC EcoRI-digested pBR322 to obtain the nadA-contg. plasmid pQAA169.  
 CC Transformation of E. coli W4546 successively with pQAB 104 and pQAA 169  
 CC produced a transformant with a 50:200 copy ratio, giving quinolinic acid  
 CC in yields which were ca. 2.6-2.8 times those obtained using W4546. The  
 CC microorganisms give high yields in a reproducible manner. Quinolinic  
 CC acid is an intermediate for pharmaceuticals and plant protection agents.  
 SQ Sequence 1724 BP; 408 A; 432 C; 470 G; 414 T;

Query Match 1.2%; Score 18; DB 1; Length 1724;  
 Best Local Similarity 100.0%; Pred. No. 7.5;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 tcattgtcaatggcgcgg 476  
 Db 976 TCATGTCATGCGCGG 959

RESULT 6  
 Q49599/C  
 ID Q49599 standard; DNA; 26 BP.  
 AC Q49599;  
 DT 27-APR-1994 (first entry)  
 DE lac gene C-terminal fragment PCR primer 2.  
 KW Polymerase chain reaction; amplification; donor plasmid;  
 KW construction; ss.  
 OS Synthetic.  
 PN EP-564965-A.  
 PD 13-OCT-1993.  
 PF 31-MAR-1993; 105303.  
 PR 31-APR-1992; EP-105973.  
 PA (NEST ) SOC PROD NESTLE SA.  
 PI Knol J, Marciset O, Mollet B;  
 DR WPI: 93-322227/41.  
 PT Integral gene expression in food-grade microorganisms - by  
 PT insertion of a promoter-less gene into an operon in front of an  
 PT essential cistron  
 PS Example 1; Page 6; 24pp; English.  
 CC The sequence is that of a PCR primer which was used in the  
 CC amplification of a ca. 900 bp long fragment, pBM37, contg. the C-terminal  
 CC end of lacS from E. coli W4546.

CC NdeI site. The PCR amplification was performed as part of the  
 CC construction of donor plasmids for the integration of the cat  
 CC gene into the lac operon.  
 SQ Sequence 26 BP; 4 A; 5 C; 3 G; 14 T;

Query Match 1.2%; Score 17; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201  
 |||||  
 DB 23 GGAGATAACATATGAA 7

RESULT 7  
 Q49600  
 ID Q49600 standard; DNA; 26 BP.  
 AC Q49600;  
 DT 27-APR-1994 (first entry)

DE lac gene C-terminal fragment PCR primer 3.  
 KW Polymerase chain reaction; amplification; donor plasmid;  
 KW construction; ss.  
 OS Synthetic.  
 PN EP-569604-A.  
 PD 13-OCT-1993.

PF 31-MAR-1993; 105303.  
 PR 07-APR-1992; EP-105973.  
 PA (NEST ) SOC PROD NESTLE SA.  
 PI Knol J, Marciset O, Mollet B;  
 DR WPI; 93-322227/41.

PT Integral gene expression in food-grade microorganisms - by  
 PT insertion of a promoter-less gene into an operon in front of an  
 PT essential cistron  
 PS Example 1; Page 6; 24pp; English.

CC The sequence is that of a PCR primer which was used in the  
 CC amplification of a ca. 700 bp long fragment, pBM39, contg. the C-terminal  
 CC end of lacS from FspI linearised pD228. The PCR amplification was  
 CC performed as part of the construction of donor plasmids for the  
 CC integration of the cat gene into the lac operon.  
 SQ Sequence 26 BP; 14 A; 3 C; 5 G; 4 T;

Query Match 1.2%; Score 17; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201  
 |||||  
 DB 4 GGAGATAACATATGAA 20

RESULT 8  
 Q51468/c  
 ID Q51468 standard; DNA; 26 BP.  
 AC Q51468;  
 DT 19-MAY-1994 (first entry)

DE Primer 2.  
 KW Polymerase chain reaction; PCR; primer; amplify; stable; integration;  
 KW homogenic; heterogenic; Streptococcus thermophilus; restriction site;  
 KW donor plasmid; artificial; milk product; yoghurt; ss.  
 OS Synthetic.  
 PN EP-569604-A.  
 PD 18-NOV-1993.

PF 07-APR-1992; 105973.  
 PR 07-APR-1992; EP-105973.  
 PA (NEST ) SOC PROD NESTLE SA.  
 PI Knol J, Marciset O, Mollet B;  
 DR WPI; 93-360915/46.

PT Stable integration of homo-, heterogenic or artificial genes into  
 PT streptococcus thermophilus genome - to produce an acidified milk  
 PT prod., pref. yoghurt  
 PS Example 2; Page 5; 9pp; English.

CC The sequences given in Q51467-72 are primers which were used in the  
 CC stable integration of homo- or heterogenic genes into the  
 CC Streptococcus thermophilus genome. These plasmids were used to  
 CC create restriction sites within donor plasmids which allowed  
 CC insertion of the required genes into the S. thermophilus genome. S.  
 CC thermophilus transformed in this way may be used in the production  
 CC of an artificial milk product, especially yoghurt. Integration is  
 CC stable and follows a strict copy number control of the host genome.  
 CC The integrated gene may be expressed without the need for direct  
 CC selection, observable phenotype or adaptation of the growth medium.  
 SQ Sequence 26 BP; 4 A; 5 C; 3 G; 14 T;

Query Match 1.2%; Score 17; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201  
 |||||  
 DB 23 GGAGATAACATATGAA 7

RESULT 9

Q51469  
 ID Q51469 standard; DNA; 26 BP.  
 AC Q51469;  
 DT 19-MAY-1994 (first entry)

DE Primer 3.

KW Polymerase chain reaction; PCR; primer; amplify; stable; integration;  
 KW homogenic; heterogenic; Streptococcus thermophilus; restriction site;  
 KW donor plasmid; artificial; milk product; yoghurt; ss.  
 OS Synthetic.  
 PN EP-569604-A.  
 PD 18-NOV-1993.

PF 07-APR-1992; 105973.

PR 07-APR-1992; EP-105973.

PA (NEST ) SOC PROD NESTLE SA.

PI Knol J, Marciset O, Mollet B;

DR WPI; 93-360915/46.

PT Stable integration of homo-, heterogenic or artificial genes into

PT streptococcus thermophilus genome - to produce an acidified milk

PT prod., pref. yoghurt

PS Example 2; Page 5; 9pp; English.

CC The sequences given in Q51467-72 are primers which were used in the

CC stable integration of homo- or heterogenic genes into the

CC Streptococcus thermophilus genome. These plasmids were used to

CC create restriction sites within donor plasmids which allowed

CC insertion of the required genes into the S. thermophilus genome. S.

CC thermophilus transformed in this way may be used in the production

CC of an artificial milk product, especially yoghurt. Integration is

CC stable and follows a strict copy number control of the host genome.

CC The integrated gene may be expressed without the need for direct

CC selection, observable phenotype or adaptation of the growth medium.

SQ Sequence 26 BP; 14 A; 3 C; 5 G; 4 T;

Query Match 1.2%; Score 17; DB 1; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ggagaataacatatgaa 201  
 |||||  
 DB 4 GGAGATAACATATGAA 20

RESULT 10

V20812

ID V20812 standard; cDNA to mRNA; 1196 BP.

AC V20812;

DT 16-JUL-1998 (first entry)

DE Human haematopoietic stem cell growth factor coding sequence.

KW Haematopoietic stem cell growth factor; SCGF; burst-promoting activity;

KW BPA; granulocyte macrophage colony stimulating activity; gene therapy;

KW GPA; haematopoietic cell disorder; bone marrow inhibition; human; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT 140. .877  
CDS /\*tag= a  
FT polyA\_signal 1097..1102  
FT /\*tag= b  
PN WO9808869-A1.  
PD 05-MAR-1998.  
PE 27-AUG-1997; J02985.  
PR 07-JUL-1997; WO-J02349.  
PR 27-AUG-1996; JP-262252.  
PR 24-MAR-1997; JP-087242.  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
PI Hiraoka A, Mio H, Sugimura A;  
DR WPI; 98-179383/16.  
DR P-PSDB; W53245.  
PT Haematopoietic stem cell growth factor - useful for, e.g. treatment  
PT and diagnosis of haematopoietic cell abnormalities and bone marrow  
PT inhibition  
PS Claim 14; Page 42-44; 85pp; Japanese.  
CC This sequence encodes the human haematopoietic stem cell growth factor  
CC (SCGF) of the invention. The polypeptide of the invention is of mammalian  
CC origin and has haematopoietic stem cell growth factor SCGF activity,  
CC including burst-promoting activity (BPA) and granulocyte macrophage  
CC colony stimulating activity (GPA). The products can be used for  
CC treatment, diagnosis and analysis of haematopoietic cell disorders and  
CC bone marrow inhibition, e.g. by cytotoxic anticancer agents such as  
CC 5-fluorouracil. The products can also be used for amplification of  
CC haematopoietic cells in vitro, e.g. for use in marrow grafting and gene  
CC therapy by insertion of SCGF gene using a suitable therapeutic vector.  
SQ Sequence 1196 BP; 260 A; 378 C; 377 G; 181 T;

Query Match 1.2%; Score 17; DB 1; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 ttctctggccacgaatgg 962  
DB 964 TTCTCTGGCCACGAATGG 980  
|||||

RESULT 11  
V00522 ID V00522 standard; DNA; 1501 BP.  
AC V00522;  
DT 20-JUL-1998 (first entry)  
DE HIV-1 breakthrough isolate clone C15.3 gp120 polypeptide DNA.  
KW HIV-1; envelope protein; gp120; MN-rgp120; vaccine; AIDS; ss.  
OS Human immunodeficiency virus type 1.  
FH Key Location/Qualifiers  
FT 1. .1491  
CDS /\*tag= a  
FT WO9801564-A1.  
PN 15-JAN-1998.  
PE 03-JUL-1997; U09690.  
PR 08-JUL-1996; US-676737.  
PA (GETH ) GENENTECH INC.  
PI Berman PW;  
DR WPI; 98-101059/09.  
DR P-PSDB; W37059.  
PT Human immunodeficiency virus gp120 sequences from vaccine  
PT breakthrough strains - useful in providing added protection against  
PT HIV not provided by original vaccine  
PS Claim 6; Page 112-114; 193pp; English.  
CC Clone C15.3 encodes a gp120 envelope polypeptide (see W37059) of  
CC a breakthrough isolate of HIV-1 obtained from an individual  
CC infected with HIV-1 through high risk activity while participating  
CC in Phase I or Phase 2 trials of MN-rgp120, a candidate recombinant  
CC gp120 HIV-1 vaccine. Nucleotide sequences (see V00517-30) for  
CC each of 7 breakthrough isolates. The envelope glycoprotein genes  
CC were obtained from proviral DNA using PCR (see V00533-38). All 7  
CC envelope glycoproteins possess sequences typical of subtype (clade)  
CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of  
CC the gp120 polypeptides from one or more of the breakthrough  
CC isolates, usually together with MN-rgp120, can provide protection  
CC against HIV strains that are sufficiently different from the  
CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer  
CC protection against those strains.  
SQ Sequence 1512 BP; 580 A; 249 C; 320 G; 363 T;

Query Match 1.2%; Score 17; DB 1; Length 1512;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggaggt 796  
DB 821 ATATAGGACCAGGAGGT 837  
|||||

RESULT 13  
X20225 ID X20225 standard; DNA; 1883 BP.

CC were obtained from proviral DNA using PCR (see V00533-38). All 7  
CC envelope glycoproteins possess sequences typical of subtype (clade)  
CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of  
CC the gp120 polypeptides from one or more of the breakthrough  
CC isolates, usually together with MN-rgp120, can provide protection  
CC against HIV strains that are sufficiently different from the  
CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer  
CC protection against those strains.  
SQ Sequence 1501 BP; 575 A; 244 C; 320 G; 362 T;

Query Match 1.2%; Score 17; DB 1; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggaggt 796  
DB 821 ATATAGGACCAGGAGGT 837  
|||||

## RESULT 12

V00521 ID V00521 standard; DNA; 1512 BP.  
AC V00521;  
DT 20-JUL-1998 (first entry)  
DE HIV-1 breakthrough isolate clone C15.2 gp120 polypeptide DNA.  
KW HIV-1; envelope protein; gp120; MN-rgp120; vaccine; AIDS; ss.  
OS Human immunodeficiency virus type 1.  
FH Key Location/Qualifiers  
FT 1. .1500  
CDS /\*tag= a  
FT WO9801564-A1.  
PN 15-JAN-1998.  
PE 03-JUL-1997; U09690.  
PR 08-JUL-1996; US-676737.  
PA (GETH ) GENENTECH INC.  
PI Berman PW;  
DR WPI; 98-101059/09.  
DR P-PSDB; W37058.  
PT Human immunodeficiency virus gp120 sequences from vaccine  
PT breakthrough strains - useful in providing added protection against  
PT HIV not provided by original vaccine  
PS Claim 6; Page 108-110; 193pp; English.  
CC Clone C15.2 encodes a gp120 envelope polypeptide (see W37058) of  
CC a breakthrough isolate of HIV-1 obtained from an individual  
CC infected with HIV-1 through high risk activity while participating  
CC in Phase I or Phase 2 trials of MN-rgp120, a candidate recombinant  
CC gp120 HIV-1 vaccine. Nucleotide sequences (see V00517-30) for  
CC each of 7 breakthrough isolates. The envelope glycoprotein genes  
CC were obtained from proviral DNA using PCR (see V00533-38). All 7  
CC envelope glycoproteins possess sequences typical of subtype (clade)  
CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of  
CC the gp120 polypeptides from one or more of the breakthrough  
CC isolates, usually together with MN-rgp120, can provide protection  
CC against HIV strains that are sufficiently different from the  
CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer  
CC protection against those strains.  
SQ Sequence 1512 BP; 580 A; 249 C; 320 G; 363 T;

Query Match 1.2%; Score 17; DB 1; Length 1512;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggaggt 796  
DB 821 ATATAGGACCAGGAGGT 837  
|||||

AC X20225;  
 DT 20-APR-1999 (first entry)  
 DE Enterococcus faecalis EF121 gene fragment.  
 KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;  
 KW detection; attenuation; antigenic; ss.  
 OS Enterococcus faecalis.  
 PN WO9805054-A2.  
 PD 12-NOV-1998.  
 PF 04-MAY-1998; U08959.  
 PR 14-NOV-1997; US-066009.  
 PR 06-MAY-1997; US-044031.  
 PR 16-MAY-1997; US-046855.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;  
 DR WPI; 99-070095/06.  
 DR P-PSDB; Y00235.  
 PT New isolated Enterococcus faecalis polynucleotides - used to develop  
 PT products for the detection of Enterococcus and for use in vaccines  
 PT for prevention or attenuation of Enterococcus infection  
 PS Claim 1; Page 219; 301pp; English.  
 CC The present sequence encodes an antigenic polypeptide fragment  
 CC isolated from Enterococcus faecalis. The present invention describes  
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.  
 CC The proteins can be used in vaccines for preventing or attenuating an  
 CC infection caused by a member of the Enterococcus genus in an animal.  
 CC They can also be used for detecting Enterococcus antibodies in a sample.  
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic  
 CC acids. Products from the present invention can also be used for  
 CC screening compounds to identify agonists and antagonists of E. faecalis  
 CC protein activity.  
 SQ Sequence 1883 BP; 666 A; 357 C; 404 G; 456 T;

Query Match 1.2%; Score 17; DB 1; Length 1883;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 416 cagtgctataaatgcc 432  
 |||||  
 DB 1005 CAGTTGCTAAATGTC 1021

RESULT 14  
 X20306  
 ID X20306 standard; DNA; 2019 BP.  
 AC X20306;  
 DT 04-MAY-1999 (first entry)  
 DE Borrelia burgdorferi polynucleotide sequence #59.  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 OS Borrelia burgdorferi.  
 PN WO9858943-A1.  
 PD 30-DEC-1998.  
 PF 18-JUN-1998; U12764.  
 PR 03-SEP-1997; US-057483.  
 PR 20-JUN-1997; US-050359.  
 PR 22-JUL-1997; US-053344.  
 PR 22-JUL-1997; US-053377.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
 PI White OR.  
 DR WPI; 99-081217/07.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 PS Claim 1; Page 1040-1041; 1128pp; English.  
 CC X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 SQ Sequence 2019 BP; 611 A; 312 C; 298 G; 798 T;

Query Match 1.2%; Score 17; DB 1; Length 2019;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 128 tgatgatcttttaatgc 144  
 |||||  
 DB 1906 TGATGATCTTTTAATGC 1922

RESULT 15  
 X28141/c  
 ID X28141 standard; DNA; 3800 BP.  
 AC X28141;  
 DT 11-JUN-1999 (first entry)  
 DE CCoAOMT promoter, PtCCoAOMT1.  
 KW CCoAOMT promoter; Poplar; caffeoyl-CoA-O-methyltransferase promoter;  
 KW PtCCoAOMT1; plant specific promoter; xylem ray cell;  
 KW lignin level alteration; ss.  
 OS Populus trichocarpa.  
 PN WO9909188-A2.  
 PD 25-FEB-1999.  
 PF 27-JUL-1998; E04988.  
 PR 13-AUG-1997; EP-202507.  
 PA (VIAA-) VIAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
 PI Boerjan W, Chen C, Meyermans H, Van Montagu M;  
 DR WPI; 99-190169/16.  
 PT Poplar caffeoyl-CoA-O-methyltransferase (CCoAOMT) promoter - to  
 PT provide transgenic plants having altered lignin level and/or  
 PT composition  
 PS Claim 1; Fig 1b; 69pp; English.  
 CC This sequence represents the Poplar (Populus trichocarpa)  
 CC caffeoyl-CoA-O-methyltransferase (CCoAOMT) promoter, PtCCoAOMT1, of the  
 CC invention. The promoter is specific for plant or tree living vessel  
 CC elements and/or xylem adjacent ray cells. The promoter is used in genetic  
 CC modification to provide transgenic plants having altered lignin level  
 CC and/or composition, which improves digestibility of forages and pulping  
 CC properties of trees.  
 SQ Sequence 3800 BP; 1225 A; 668 C; 701 G; 1206 T;

Query Match 1.2%; Score 17; DB 1; Length 3800;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 924 atttgatcaagaactcg 940  
 |||||  
 DB 2765 ATTTGATCAAGAACTGG 2749

Search completed: September 9, 2000, 22:03:12  
 Job time: 2194 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 21:11:03 ; Search time 51.17 Seconds  
(without alignments)  
3876.402 Million cell updates/sec

Title: us-09-214-679-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtgcc.....aatgcaattcatttgatcc 1442

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 243080 seqs, 68777915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCFUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Résult No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| c 1        | 17    | 1.2         | 26     | 1     | US-08-039-866-2    |
| 2          | 17    | 1.2         | 26     | 1     | US-08-039-866-3    |
| 3          | 16    | 1.1         | 75     | 1     | US-07-731-157A-38  |
| 4          | 16    | 1.1         | 75     | 4     | US-08-541-780-38   |
| 5          | 16    | 1.1         | 109    | 3     | US-08-417-210A-122 |
| 6          | 16    | 1.1         | 109    | 3     | US-08-417-210A-145 |
| 7          | 16    | 1.1         | 475    | 3     | US-08-967-101-32   |
| 8          | 16    | 1.1         | 475    | 4     | US-08-592-541-32   |
| 9          | 16    | 1.1         | 1350   | 1     | US-08-665-435A-1   |
| 10         | 16    | 1.1         | 1350   | 1     | US-08-665-435A-3   |
| 11         | 16    | 1.1         | 1350   | 3     | US-08-843-309-3    |
| 12         | 16    | 1.1         | 1353   | 3     | US-08-843-309-1    |
| 13         | 16    | 1.1         | 1532   | 3     | US-08-037-816A-15  |
| 14         | 16    | 1.1         | 1532   | 3     | US-08-037-816A-27  |
| 15         | 16    | 1.1         | 1532   | 3     | US-08-530-146-15   |
| 16         | 16    | 1.1         | 1532   | 3     | US-08-530-146-27   |
| 17         | 16    | 1.1         | 2019   | 5     | US-08-926-636-1    |
| 18         | 16    | 1.1         | 2020   | 3     | US-08-417-210A-135 |
| 19         | 16    | 1.1         | 2028   | 3     | US-08-417-210A-138 |
| 20         | 16    | 1.1         | 2060   | 3     | US-08-417-210A-141 |
| 21         | 16    | 1.1         | 2161   | 1     | US-08-399-646-11   |
| 22         | 16    | 1.1         | 2161   | 2     | US-08-607-321-11   |
| 23         | 16    | 1.1         | 2161   | 1     | US-08-961-240-11   |
| 24         | 16    | 1.1         | 2161   | 3     | US-08-605-501-11   |
| 25         | 16    | 1.1         | 2163   | 1     | US-07-731-157A-1   |
| c 26       | 16    | 1.1         | 2163   | 4     | US-08-541-780-1    |

27 16 1.1 2193 4 US-08-934-481-1 Sequence 1, Appli  
28 16 1.1 2552 3 US-08-448-603A-27 Sequence 27, Appl  
29 16 1.1 2552 5 US-09-134-075-27 Sequence 27, Appl  
30 16 1.1 2571 1 US-08-254-358-3 Sequence 3, Appli  
31 16 1.1 2571 2 US-08-475-391-3 Sequence 3, Appli  
32 16 1.1 2571 3 US-08-709-609-3 Sequence 3, Appli  
33 16 1.1 2571 6 PCT-US95-07178-3 Sequence 3, Appli  
34 16 1.1 2995 2 US-08-592-126-85 Sequence 85, Appl  
35 16 1.1 3249 1 US-08-106-493A-1 Sequence 1, Appli  
36 16 1.1 3249 1 US-08-429-264-1 Sequence 1, Appli  
37 16 1.1 3807 3 US-08-417-210A-78 Sequence 78, Appl  
38 16 1.1 4853 2 US-08-832-883-1 Sequence 1, Appli  
39 16 1.1 4853 3 US-08-832-877-1 Sequence 1, Appli  
40 16 1.1 7797 4 US-08-816-1558-7 Sequence 7, Appli  
41 16 1.1 8535 5 US-08-716-351A-1 Sequence 1, Appli  
42 16 1.1 8654 2 US-08-920-812-6 Sequence 6, Appli  
43 16 1.1 8654 2 US-08-920-827-6 Sequence 6, Appli  
44 16 1.1 8654 2 US-08-921-177-6 Sequence 6, Appli  
45 16 1.1 8654 2 US-08-362-577C-6 Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-08-039-866-2/c  
; Sequence 2, Application US/08039866  
; Patent No. 5491079  
; GENERAL INFORMATION:  
; APPLICANT: KNOL, Jan  
; APPLICANT: MARCISSET, Olivier  
; APPLICANT: MOLLET, Beat  
; TITLE OF INVENTION: Integrative gene-expression in  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: VOGT & O'DONNELL - STEPHEN M. HARACZ  
; STREET: 707 Westchester Avenue  
; CITY: WHITE PLAINS  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10604  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/039,866  
; FILING DATE: 30-MAR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92105973  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HARACZ, STEPHEN M.  
; REGISTRATION NUMBER: 33397  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914) 328-0055  
; TELEFAX: (914) 328-0060  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-039-866-2

Query Match 1.2%; Score 17; DB 1; Length 26;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/731.157  
; FILING DATE:  
; APPLICATION NUMBER: EP 90200962  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RAE-VENTER PH.D., BARBARA  
; REGISTRATION NUMBER: 32,750  
; REFERENCE/DOCKET NUMBER: GBRO-027/0005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; TELEX: 380816 COOLEY PA  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 75 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (synthetic)  
; HYPOTHETICAL: NO  
US-08-541-780-38

Query Match 1.1%; Score 16; DB 4; Length 75;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 838 tattgtgatgcccat 853  
|||||  
DB 33 TATTGTGATGCCCAT 48

RESULT 5  
US-08-417-210A-122  
; Sequence 122, Application US/08417210A  
; Patent No. 5863542  
; GENERAL INFORMATION:  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: TARTAGLIA, JAMES  
; APPLICANT: COX, WILLIAM I.  
; TITLE OF INVENTION: IMMUNODEFICIENCY RECOMBINANT POXVIRUS  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
; STREET: 530 FIFTH AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417.210A  
; FILING DATE: 05-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOWALSKI, THOMAS J.  
; REGISTRATION NUMBER: 32,147  
; REFERENCE/DOCKET NUMBER: 454310-2690  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-840-3333  
; INFORMATION FOR SEQ ID NO: 122:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 109 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-417-210A-122

Query Match 1.1%; Score 16; DB 3; Length 109;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795  
|||||  
DB 53 ATATAGGACCAGGAG 68

RESULT 6  
US-08-417-210A-145  
; Sequence 145, Application US/08417210A  
; Patent No. 5863542  
; GENERAL INFORMATION:  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: TARTAGLIA, JAMES  
; APPLICANT: COX, WILLIAM I.  
; TITLE OF INVENTION: IMMUNODEFICIENCY RECOMBINANT POXVIRUS  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
; STREET: 530 FIFTH AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417.210A  
; FILING DATE: 05-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOWALSKI, THOMAS J.  
; REGISTRATION NUMBER: 32,147  
; REFERENCE/DOCKET NUMBER: 454310-2690  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-840-3333  
; INFORMATION FOR SEQ ID NO: 145:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 109 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-417-210A-145

Query Match 1.1%; Score 16; DB 3; Length 109;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795  
|||||  
DB 53 ATATAGGACCAGGAG 68

RESULT 7  
US-08-967-101-32  
; Sequence 32, Application US/08967101  
; Patent No. 5840540  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183

us-09-214-679-1.olig.rni

Tue Sep 12 08:47:28 2000

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
;; STREET: High Street Tower - 125 High Street  
;; CITY: Boston  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02110  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/967,101  
;; FILING DATE: 10-NOV-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/592,541  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Pitcher, Edmund R.  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 248-7000  
;; TELEFAX: (617) 248-7100  
;; INFORMATION FOR SEQ ID NO: 32:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 475 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-967-101-32

Query Match 1.1%; Score 16; DB 3; Length 475;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 972 aaaatattatgagtat 987  
|||||

Db 37 AAAATATTATGAGTAT 52

RESULT 8  
US-08-592-541-32  
; Sequence 32, Application US/08592541  
; Patent No. 5986054  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 183  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: High Street Tower - 125 High Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,541  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pitcher, Edmund R.  
; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617) 248-7000  
;; TELEFAX: (617) 248-7100  
;; INFORMATION FOR SEQ ID NO: 32:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 475 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-592-541-32

Query Match 1.1%; Score 16; DB 4; Length 475;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 972 aaaatattatgagtat 987  
|||||

Db 37 AAAATATTATGAGTAT 52

RESULT 9  
US-08-665-435A-1  
; Sequence 1, Application US/08665435A  
; Patent No. 5681694  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul  
; APPLICANT: Peery, Robert  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Wu, Chyun-Yeh Earnest  
; TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus  
; TITLE OF INVENTION: pneumoniae  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: US  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,435A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-9900  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; TELEFAX: 317-276-3861  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1350 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA(genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1350  
; US-08-665-435A-1

Query Match 1.1%; Score 16; DB 1; Length 1350;  
Best Local Similarity 100.0%; Pred. No. 45;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1117 gctgggcaacatggctc 1132  
|||||  
Db 409 CTGGGGCAACATGGTC 424

## RESULT 10

US-08-665-435A-3  
; Sequence 3, Application US/08665435A  
; Patent No. 5681694  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul  
; APPLICANT: Peery, Robert  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Wu, Chyun-Yeh Earnest  
; TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus  
; TITLE OF INVENTION: pneumoniae  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: US  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,435A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-9900  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; TELEFAX: 317-276-3861  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1350 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-665-435A-3

Query Match 1.1%; Score 16; DB 1; Length 1350;  
Best Local Similarity 81.2%; Pred. No. 45;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1117 gctgggcaacatggctc 1132  
|||||  
Db 409 GCUGGGCAACAUUGGUC 424

## RESULT 11

US-08-843-309-3  
; Sequence 3, Application US/08843309  
; Patent No. 5834270  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul  
; APPLICANT: Peery, Robert  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Wu, Chyun-Yeh Earnest  
; TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae

; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: US  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/843,309  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-9900  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; TELEFAX: 317-276-3861  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1350 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-843-309-3

Query Match 1.1%; Score 16; DB 3; Length 1350;  
Best Local Similarity 81.2%; Pred. No. 45;  
Matches 13; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1117 gctgggcaacatggctc 1132  
|||||  
Db 409 GCUGGGCAACAUUGGUC 424

## RESULT 12

US-08-843-309-1  
; Sequence 1, Application US/08843309  
; Patent No. 5834270  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul  
; APPLICANT: Peery, Robert  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Wu, Chyun-Yeh Earnest  
; TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: US  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/843,309  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

Tue Sep 12 08:47:28 2000

```

; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1353
; US-08-843-309-1

Query Match 1.1%; Score 16; DB 3; Length 1353;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1117 gctgggcaacatggtc 1132
|||||
DB 409 GCTGGGCAACATGGTC 424

RESULT 13
US-08-037-816A-15
; Sequence 15, Application US/08037816A
; Patent No. 5869624
; GENERAL INFORMATION:
; APPLICANT: Hasel, Karl W.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
; TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,816A
; FILING DATE: 26-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOPUI
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1522
; OTHER INFORMATION:
; US-08-037-816A-27

Query Match 1.1%; Score 16; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795
|||||
DB 932 ATATAGGACCAGGAG 947
```

```

; NAME/KEY: CDS
; LOCATION: 1..1522
; OTHER INFORMATION:
; US-08-037-816A-15

Query Match 1.1%; Score 16; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795
|||||
DB 932 ATATAGGACCAGGAG 947

RESULT 14
US-08-037-816A-27
; Sequence 27, Application US/08037816A
; Patent No. 5869624
; GENERAL INFORMATION:
; APPLICANT: Hasel, Karl W.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
; TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,816A
; FILING DATE: 26-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOPUI
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1522
; OTHER INFORMATION:
; US-08-037-816A-27

Query Match 1.1%; Score 16; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 atataggaccaggag 795
|||||
DB 932 ATATAGGACCAGGAG 947
```

RESULT 15

US-08-530-146-15  
 ; Sequence 15, Application US/08530146  
 ; Patent No. 5886163  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hasel, Karl W.  
 ; APPLICANT: Maddon, Paul J.  
 ; TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
 ; TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES  
 ; TITLE OF INVENTION: THEREOF  
 ; NUMBER OF SEQUENCES: 29  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham  
 ; STREET: 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.24  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/530,146  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/037,816  
 ; FILING DATE: 26-MAR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 41190/JPW/AJM  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 977-9550  
 ; TELEFAX: (212) 664-0525  
 ; TELEX: 422523 COOPUI  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; - SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1532 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; - MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1522  
 ; OTHER INFORMATION:  
 ; US-08-530-146-15

Query Match 1.1%; Score 16; DB 3; Length 1532;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 780 atattagaccaggag 795  
 |||  
 Db 932 ATATTAGCAGGAG 947

Search completed: September 9, 2000, 22:01:48  
 Job time: 3045 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:45:57 ; Search time 585.75 Seconds  
(without alignments)  
10856.120 Million cell updates/sec

Title: US-09-214-679-1  
Perfect score: 1442  
Sequence: 1 cccgggaactccatgtggcc.....aatgcaatcatttgatcc 1442

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5247842 seqs, 2204914090 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: gb\_gss7:\*  
104: gb\_gss8:\*  
105: gb\_gss9:\*  
106: em\_gss5:\*  
107: em\_gss6:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: em\_gss9:\*  
111: em\_gss10:\*  
112: em\_gss11:\*  
113: gb\_gss10:\*  
114: gb\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*

117: gb\_gss13:\*  
 118: gb\_gss14:\*  
 119: gb\_gss15:\*  
 120: gb\_gss16:\*  
 121: gb\_gss17:\*  
 122: gb\_gss18:\*  
 123: gb\_gss19:\*  
 124: em\_gss13:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query Match | Length | ID  | Description         |
|------------|-------------|--------|-----|---------------------|
| C 1        | 21          | 409    | 120 | B44876 HS-1060-Al-  |
| C 2        | 21          | 454    | 91  | W66727 me26d06.r1   |
| C 3        | 20          | 1.4    | 71  | AW357770 41461 MAR  |
| C 4        | 20          | 1.4    | 71  | AW357770 41461 MAR  |
| C 5        | 20          | 1.4    | 245 | AW356799 39389 MAR  |
| C 6        | 20          | 1.4    | 261 | AW326093 18169 MAR  |
| C 7        | 20          | 1.4    | 279 | AW357182 40066 MAR  |
| C 8        | 20          | 1.4    | 329 | AW357182 40066 MAR  |
| C 9        | 20          | 1.4    | 382 | N74414 za3c02.s1    |
| C 10       | 20          | 1.4    | 420 | W87700 zh65b11.s1   |
| C 11       | 20          | 1.4    | 550 | AQ529126 RPCT-11-3  |
| C 12       | 20          | 1.4    | 572 | AQ529126 RPCT-11-3  |
| C 13       | 20          | 1.4    | 595 | C79644 C79644 Mous  |
| C 14       | 19          | 1.3    | 330 | AQ597255 HS_5300_B  |
| C 15       | 19          | 1.3    | 360 | AI146143 UI-R-BTO-  |
| C 16       | 19          | 1.3    | 376 | AV193600 AV193600   |
| C 17       | 19          | 1.3    | 378 | C68576 C68576 Yuj1  |
| C 18       | 19          | 1.3    | 429 | AV203736 AV203736   |
| C 19       | 19          | 1.3    | 440 | M89105 CEL16D1 Chr  |
| C 20       | 19          | 1.3    | 456 | AA799513 EST199010  |
| C 21       | 19          | 1.3    | 484 | AI555490 UI-R-C2P-  |
| C 22       | 19          | 1.3    | 509 | B36341 HS-1039-Al-  |
| C 23       | 19          | 1.3    | 527 | H76674 17105 Lambd  |
| C 24       | 19          | 1.3    | 557 | AQ036697 CIT-HSP-2  |
| C 25       | 19          | 1.3    | 563 | AW435445 UI-R-BJOP  |
| C 26       | 19          | 1.3    | 592 | AI683139 tw49a08.x  |
| C 27       | 19          | 1.3    | 636 | AZ037112 RPCT-23-2  |
| C 28       | 19          | 1.3    | 743 | AQ0803720 HS_3175_A |
| C 29       | 18          | 1.2    | 227 | AW081978 xb58c10.x  |
| C 30       | 18          | 1.2    | 300 | AQ079435 CIT-HSP-2  |
| C 31       | 18          | 1.2    | 300 | C34626 C34626 Yuj1  |
| C 32       | 18          | 1.2    | 301 | C37085 C37085 Yuj1  |
| C 33       | 18          | 1.2    | 311 | AI112615 UI-R-Y0-m  |
| C 34       | 18          | 1.2    | 339 | AI917291 ts96f12.x  |
| C 35       | 18          | 1.2    | 387 | C15144 C15144 Clon  |
| C 36       | 18          | 1.2    | 387 | C82966 C82966 rabb  |
| C 37       | 18          | 1.2    | 391 | C83822 C83822 rabb  |
| C 38       | 18          | 1.2    | 396 | D86807 D86807 Rac   |
| C 39       | 18          | 1.2    | 403 | AW38993 gp23c10.y   |
| C 40       | 18          | 1.2    | 410 | H65770 Yr74b09.r1   |
| C 41       | 18          | 1.2    | 414 | AA487512 aa95d01.s  |
| C 42       | 18          | 1.2    | 422 | AI144901 UI-R-BTO-  |
| C 43       | 18          | 1.2    | 424 | AW335528 ra21c12.y  |
| C 44       | 18          | 1.2    | 429 | AW335528 ra21c12.y  |
| C 45       | 18          | 1.2    | 438 | AA652960 ns71d08.s  |
|            |             |        |     | AQ022853 HS_2185_A  |

## ALIGNMENTS

RESULT 1  
 B44876/c  
 LOCUS  
 DEFINITION  
 B44876 409 bp DNA  
 HS-1060-Al-C06-MR.abi CIT Human Genomic Sperm Library C Homo  
 sapiens genomic clone Plate-Ct 782 Col-11 Row-M, genomic

## ACCESSION

VERSION B44876.1 GI:2549710

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 409)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.

TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors

## JOURNAL

COMMENT Unpublished (1997)

On Dec 15, 1999 this sequence version replaced gi:4213071.

Contact: Mahairas GG, Zackrone KD, Hood L

University of Washington

Seattle, WA 98195, USA

Tel: (206) 616-8744

Fax: (206) 685-7301

Email: kzackron@u.washington.edu

Sequence Tagged Connector

Plate: Ct 782 row: M column: 11

Class: BAC ends

High quality sequence stop: 409.

Location/Qualifiers

1..409

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="plate=Ct 782 Col=11 Row=M"

/clone\_lib="CIT Human Genomic Sperm Library C"

/sex="M"

/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 110 a 82 c 90 g 127 t

ORIGIN

Query Match 1.5%: Score 21; DB 120; Length 409;

Best Local Similarity 100.0%; Pred. No. 0.85; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 atgaatgggttggaagaatcc 217

|||||

Db 179 ATGAATGGTTGGGAATCC 159

|||||

RESULT 2

W66727/c

LOCUS

me26d06.r1 Soares mouse embryo NbMe13.5 14.5 Mus musculus CDNA

clone IMAGE:388619 5', mRNA sequence.

W66727

W66727.1 GI:1375662

EST.

house mouse.

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 454)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

On Jan 7, 1998 this sequence version replaced gi:636815.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:240451

Seq primer: ETPrimer

High quality sequence stop: 341.

Location/Qualifiers

#### FEATURES

source

1. 454

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:38619"

/clone\_lib="Soares mouse embryo NDME13.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5,

TGTTACCAATCTGAAGTGGGCGCGCGGAAATTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos (total RNA provided by Minoru Ko, Wayne

State Univ., from 2 1; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT7T3 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M.Fatima Bonaldo."

99 a 117 c 104 g 134 t

#### BASE COUNT

ORIGIN

#### Query Match

Best Local Similarity 1.5%; Score 21; DB 91; Length 454;

.Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 904 aatcaccaccatcaagtcga 924

|||||

Db 49 AATCACCACCAATCAAGTCGA 29

#### RESULT

3

AW357770

LOCUS

41461 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

AW357770

ACCESSION

AW357770.1

VERSION

AW357770.1

KEYWORDS

EST.

SOURCE

Bos taurus.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 240)

Smith,T.P.I., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,

Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and

Keele,J.W.

Design and use of four pooled tissue normalized cDNA libraries for

EST discovery in cattle

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:5676822.

Contact: Smith TPL

PO Box 166, Clay Center, NE 68933-0166, USA

USDA, ARS, US Meat Animal Research Center

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithemall.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 20

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCCACTCAGCAGC

Plate: 18 row: N column: 4

Seq primer: ATTTAGTGACACTATAG.

#### FEATURES

source

1. 240

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 3BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;

Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal

longissimus muscle."

33 a 75 c 59 g 73 t

BASE COUNT

ORIGIN

#### Query Match

Best Local Similarity 1.4%; Score 20; DB 71; Length 240;

.Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 420 tgctaaaaatgcctttctc 439

|||||

Db 15 TGCTAAATGCCCTTCTC 34

#### RESULT

4

AW356799

LOCUS

39389 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

AW356799

ACCESSION

AW356799.1

VERSION

AW356799.1

KEYWORDS

EST.

SOURCE

Bos taurus.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 245)

Smith,T.P.I., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,

Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and

Keele,J.W.

Design and use of four pooled tissue normalized cDNA libraries for

EST discovery in cattle

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:6575823.

Contact: Smith TPL

PO Box 166, Clay Center, NE 68933-0166, USA

USDA, ARS, US Meat Animal Research Center

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithemall.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 20

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCCACTCAGCAGC

Plate: 15 row: H column: 16

Seq primer: ATTTAGTGACACTATAG.

Location/Qualifiers

1. 245

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 3BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;

Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle." 64 t

BASE COUNT 47 a 87 c 47 g 64 t

ORIGIN

Query Match 1.4%; Score 20; DB 71; Length 245;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 420 tgctaaaaatgccctttctc 439  
|||||

Db 147 TGCTAAAAATGCCCTTTCTC 166

RESULT 5

AW326093 261 bp mRNA EST 27-JAN-2000

LOCUS

DEFINITION 18169 MARC IBOV Bos taurus CDNA 5', mRNA sequence.

ACCESSION AW326093

VERSION AW326093.1 GI:6762014

KEYWORDS

SOURCE EST.

ORGANISM Bos taurus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.

AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and Keel,J.W.

TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle

JOURNAL Unpublished (2000)

COMMENT On Jan 6, 2000 this sequence version replaced gi:6675590.

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 20 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 5 row: 1 column: 23

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..261

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC IBOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 44 a 97 c 55 g 65 t

ORIGIN

Query Match 1.4%; Score 20; DB 71; Length 261;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 420 tgctaaaaatgccctttctc 439  
|||||

Db 128 TGCTAAAAATGCCCTTTCTC 147

RESULT 6

AW357182 279 bp mRNA EST 03-FEB-2000

LOCUS

DEFINITION 40056 MARC 2BOV Bos taurus CDNA 5', mRNA sequence.

ACCESSION AW357182

VERSION AW357182.1 GI:6861188

KEYWORDS

SOURCE EST.

ORGANISM Bos taurus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.

AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and Keel,J.W.

TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle

JOURNAL Unpublished (2000)

COMMENT Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 20 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 23 row: 1 column: 15

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..279

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 2BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."

BASE COUNT 50 a 96 c 54 g 79 t

ORIGIN

Query Match 1.4%; Score 20; DB 71; Length 279;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 420 tgctaaaaatgccctttctc 439  
|||||

Db 145 TGCTAAAAATGCCCTTTCTC 164

RESULT 7

N74414 329 bp mRNA EST 19-MAR-1996

LOCUS

DEFINITION za53c02.sl Soares fetal liver spleen 1NPLS Homo sapiens CDNA clone IMAGE:296258 3', mRNA sequence.

ACCESSION N74414

VERSION N74414.1 GI:1231699

KEYWORDS

SOURCE EST.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; 1 (bases 1 to 329)

AUTHORS Hillier,L., Clark,N., Dubouque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and

```

Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On Dec 20, 1995 this sequence version replaced gi:1133302.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 308.
Location/Qualifiers
1. 329
/organism="Homo sapiens"
/db_xref="GDB:1241179"
/db_xref="taxon:9606"
/clone="IMAGE:296258"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
111 a 81 c 73 g 63 t 1 others
BASE COUNT
ORIGIN

Query Match 1.4%; Score 20; DB 87; Length 329;
Best Local Similarity 100.0%; Pred. No. 3.2;
-Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 gctatcaattcggaacagga 405
|||||
Db 30 GCTATCAATTCGGAACAGGA 49

RESULT 8
W87700 382 bp mRNA EST 02-FEB-1997
DEFINITION zh65bil.s1 Soares_fetal_liver_spleen_INFLS_s1 Homo sapiens cDNA
clone IMAGE:416925 3', mRNA sequence.
ACCESSION W87700
VERSION W87700.1 GI:1401754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 382)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissee, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
9704478
On Jan 5, 1998 this sequence version replaced gi:1327905.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```

```

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 620 Std Error: 0.00
Seq primer: mob.REGA+ET.
Location/Qualifiers
1. 382
/organism="Homo sapiens"
/db_xref="GDB:1325395"
/db_xref="taxon:9606"
/clone="IMAGE:416925"
/clone_lib="Soares_fetal_liver_spleen_INFLS_s1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
119 a 99 c 93 g 68 t 3 others
BASE COUNT
ORIGIN

Query Match 1.4%; Score 20; DB 91; Length 382;
Best Local Similarity 100.0%; Pred. No. 3.2;
-Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 gctatcaattcggaacagga 405
|||||
Db 33 GCTATCAATTCGGAACAGGA 52

RESULT 9
AU006865 420 bp mRNA EST 31-JUL-1998
LOCUS AU006865 Schizosaccharomyces pombe late log phase cDNA
DEFINITION Schizosaccharomyces pombe cDNA clone spc01051, mRNA sequence.
ACCESSION AU006865
VERSION AU006865.1 GI:3343323
KEYWORDS EST.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 420)
AUTHORS Morimyo, M. and Mita, K.
TITLE Identification of expressed sequence tags of Schizosaccharomyces
pombe
JOURNAL Unpublished (1998)
COMMENT On May 14, 1999 this sequence version replaced gi:4827815.
Contact: Mitsuoki Morimyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: morimyo@nirs.go.jp.
Location/Qualifiers
1. 420
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc01051"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of

```

Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mp19 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, <http://www.nirs.go.jp>)"

BASE COUNT 156 a 62 c 78 g 124 t  
ORIGIN

Query Match 1.4%; Score 20; DB 47; Length 420;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 960 tggagaatgccgaataatt 979  
|||||  
Db 331 TGGAGATGCCGAAATATT 350

RESULT 10  
LOCUS AQ529126 550 bp DNA GSS 18-MAY-1999  
DEFINITION RPCI-11-368P15-TV RPCI-11 Homo sapiens genomic clone RPCI-11-368P15, genomic survey sequence.  
ACCESSION AQ529126

VERSION AQ529126.1 GI:4841239  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 550)  
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT On Sep 10, 1998 this sequence version replaced gi:3555790.

Other\_GSSs: RPCI-11-368P15.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (info@resgen.com). BAC end search page: [http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..550  
/organism="Homo sapiens"  
/db\_xref="GDB:7641302"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-368P15"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT 174 a 80 c 109 g 187 t  
ORIGIN

Query Match 1.4%; Score 20; DB 103; Length 550;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1178 gtttagtaggaataaataac 1197  
|||||

Db 384 GTTTAGTAGGAATAACTAAC 403

RESULT 11  
C79644

LOCUS C79644 572 bp mRNA EST 26-JUN-1998  
DEFINITION J0069G09 3', mRNA sequence.

ACCESSION C79644  
VERSION C79644.1 GI:2519974  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 572)  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Ko,M.S.H., Kitchen,J.R., Wang,X., Wang,X., Threat,T.A., Sun,T., Grahovac,M.J., Mason,S., Lim,M.K., Paonessa,P.D., Sauls,A.D. and Doi,H.

TITLE Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst (The ERATO/Doi Project at Wayne State University)  
JOURNAL Unpublished (1997)  
COMMENT Contact: Hirofumi Doi

Doi Biosymmetry Project, ERATO  
Japan Science and Technology Corporation (JST)  
WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan  
Email: hdbioa.jst.go.jp

FEATURES  
source Location/Qualifiers  
1..572

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="J0069G09"  
/clone\_lib="Mouse 3.5-dpc blastocyst cDNA"  
/tissue\_type="blastocyst"  
/dev\_stage="3.5-dpc"  
BASE COUNT 149 a 136 c 113 g 167 t 7 others  
ORIGIN

Query Match 1.4%; Score 20; DB 81; Length 572;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 actggcagcttctctggcca 955  
|||||

Db 354 ACTGGCAGCTTCTCTGGCCA 373

RESULT 12

LOCUS AQ597255 595 bp DNA GSS 08-JUN-1999  
DEFINITION HS\_5300.B1.A11.T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-876 Col-21 Row-B, genomic survey sequence.  
ACCESSION AQ597255

VERSION AQ597255.1 GI:5028467  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 595)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequencing-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Resear h Genetics ([info@resgen.com](http://info@resgen.com)). BAC end Web Server: <http://www.husc.washington.edu>  
Plate: 876 row: B column: 21  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 595.

## FEATURES

source

1. .595  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-876 Col=21 Row=B"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"  
173 a 96 c 116 g 193 t 17 others

BASE COUNT  
ORIGIN

Query Match 1.4%; Score 20; DB 104; Length 595;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1178 gtttagtaggaataactaac 1197  
|||||  
Db 421 GTTTAGTAGGAATAACTAAC 440

## RESULT 13

A1146143

LOCUS A1146143 330 bp mRNA EST 05-JUL-1999  
DEFINITION UI-R-BT0-qh-h-04-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone  
UI-R-BT0-qh-h-04-0-UI 3', mRNA sequence.  
ACCESSION A1146143  
VERSION A1146143.1 GI:3667942  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 330)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477

On Apr 7, 1998 this sequence version replaced gi:3035792.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.  
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the

I.M.A.G.E. Consortium at LNL ([info@image.llnl.gov](mailto:info@image.llnl.gov)). IMAGE  
ID=1788816 The following repetitive elements were found in this  
cDNA sequence: 55-85, >GC\_rich#Low\_complexity  
Seq primer: M13 Forward  
POLYA=No.

## FEATURES

source

1. .330  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BT0-qh-h-04-0-UI"  
/clone\_lib="UI-R-BT0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker: Site\_1: Not I; Site\_2: Eco RI; This library  
(UI-R-BT0) consists of a mixture of individually tagged  
normalized libraries constructed from rat hippocampus,  
thalamus, mid-brain, medulla, corpus striatum, cerebral  
cortex and testis. The tag used to identify the source  
tissue is a string of 3-6 nucleotides present between the  
Not I site and the oligo-dT track which allows  
identification of the library of origin of a clone within  
the mixture. This library was then subtracted using a  
driver consisting of a mixture of all clones from UI-R-A0,  
UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and  
UI-R-C2p."  
42 a 128 c 118 g 42 t

BASE COUNT  
ORIGIN

Query Match 1.3%; Score 19; DB 35; Length 330;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1001 ccgctggaggatgcgcgc 1019  
|||||  
Db 276 CCGCTGGAGGATGCGCGC 294

## RESULT 14

AV193600/c

LOCUS AV193600 360 bp mRNA EST 22-JUL-1999  
DEFINITION AV193600 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite  
embryo Caenorhabditis elegans cDNA clone yk619c5 5', mRNA sequence.  
ACCESSION AV193600  
VERSION AV193600.1 GI:5575752  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 360)  
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,  
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,  
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and  
Nomoto,H.  
Expressed genes in C.elegans  
Unpublished (1999)  
On Jul 9, 1999 this sequence version replaced gi:5434899.  
Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.  
Location/Qualifiers  
1. .360  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk619c5"

Tue Sep 12 08:47:29 2000

/clone\_lib="Yuji Kohara unpublished cdna:Strain N2  
hermaphrodite embryo"  
/sex="hermaphrodite"  
/dev\_stage="embryo"  
BASE COUNT 129 a 72 c 50 g 109 t  
ORIGIN

Query Match 1.3%; Score 19; DB 51; Length 360;  
Best Local Similarity 100.0%; Pred. NO. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 atatgaatggttgaga 213  
|||||  
DB 112 ATATGAATGTTGGAAGA 94

RESULT 15  
C68576 376 bp mRNA EST 23-SEP-1997  
LOCUS C68576 Yuji Kohara unpublished cdna Caenorhabditis elegans cdna  
DEFINITION clone yk305f6 5', mRNA sequence.  
C68576  
ACCESSION C68576.1 GI:2429932  
VERSION EST.  
KEYWORDS Caenorhabditis elegans.  
SOURCE Caenorhabditis elegans  
ORGANISM Caenorhabditis elegans  
REFERENCE  
1 (bases 1 to 376)  
Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,  
Sano,M., Miyata,A. and Nishigaki,A.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
Expression map of the C.elegans genome  
Unpublished (1996)  
On Jun 15, 1998 this sequence version replaced gi:3224348.  
TITLE  
JOURNAL  
COMMENT  
Contact: Yuji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.

FEATURES  
source  
1..376  
Location/Qualifiers  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(el489)"  
/db\_xref="taxon:6239"  
/clone="yk305f6"  
/clone\_lib="Yuji Kohara unpublished cdna"  
/sex="hermaphrodite, male"  
/tissue\_type="whole animal"  
/dev\_stage="varied"  
BASE COUNT 109 a 75 c 84 g 108 t  
ORIGIN

Query Match 1.3%; Score 19; DB 81; Length 376;  
Best Local Similarity 100.0%; Pred. NO. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 aacatgaatggttgga 210  
|||||  
DB 138 AACATGAATGTTGGA 156

Search completed: September 9, 2000, 21:38:20  
Job time: 3143 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:17:02 ; Search time 1336.1 Seconds  
(without alignments)  
1926.152 Million cell updates/sec

Title: US-09-214-679-1  
Perfect score: 1442  
Sequence: 1 cccgggaactccatgtggcc.....aatgcaattcatttgatcc 1442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: em\_fun.\*
- 17: em\_hum1.\*
- 18: em\_hum2.\*
- 19: em\_in.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_sy.\*
- 29: em\_un.\*
- 30: em\_vi.\*
- 31: gb\_htg1.\*
- 32: gb\_htg2.\*
- 33: gb\_in1.\*
- 34: gb\_in2.\*
- 35: em\_ba1.\*
- 36: em\_ba2.\*
- 37: em\_hum3.\*
- 38: em\_hum4.\*
- 39: gb\_pr4.\*
- 40: gb\_htg3.\*
- 41: gb\_htg4.\*
- 42: gb\_htg5.\*
- 43: gb\_htg6.\*

- 44: gb\_htg7.\*
- 45: em\_htg1.\*
- 46: em\_htg2.\*
- 47: em\_htg3.\*
- 48: em\_hum5.\*
- 49: gb\_pl3.\*
- 50: gb\_pr5.\*
- 51: gb\_htg8.\*
- 52: gb\_htg9.\*
- 53: gb\_htg10.\*
- 54: gb\_htg11.\*
- 55: gb\_htg12.\*
- 56: gb\_htg13.\*
- 57: gb\_htg14.\*
- 58: gb\_in3.\*
- 59: gb\_htg15.\*
- 60: gb\_htg16.\*
- 61: gb\_htg17.\*
- 62: em\_htg4.\*
- 63: em\_htg5.\*
- 64: em\_htg6.\*
- 65: em\_htg7.\*
- 66: em\_hum6.\*
- 67: gb\_htg18.\*
- 68: gb\_htg19.\*
- 69: gb\_htg20.\*
- 70: gb\_htg21.\*
- 71: gb\_htg22.\*
- 72: gb\_htg23.\*
- 73: gb\_htg24.\*
- 74: gb\_htg25.\*
- 75: gb\_htg26.\*
- 76: gb\_htg27.\*
- 77: gb\_htg28.\*
- 78: gb\_htg29.\*
- 79: gb\_htg30.\*
- 80: gb\_htg31.\*
- 81: gb\_vil.\*
- 82: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query % |       |        | DB ID       | Description         |
|------------|---------|-------|--------|-------------|---------------------|
|            | Score   | Match | Length |             |                     |
| 1          | 1442    | 100.0 | 1442   | 5 A69475    | A69475 Sequence 1   |
| 2          | 1442    | 100.0 | 1442   | 5 A72152    | A72152 Sequence 1   |
| C 3        | 45.8    | 3.2   | 167000 | 1 AP000059  | AP000059 Aeropyrum  |
| C 4        | 42.8    | 3.0   | 35346  | 8 SPAC869   | AL132779 S.pombe c  |
| C 5        | 41.2    | 2.9   | 307120 | 1 CNSPAX03  | AJ248285 Pyrococcus |
| C 6        | 41.2    | 2.9   | 349061 | 2 NMA222491 | AL162753 Neisseria  |
| C 7        | 39.6    | 2.7   | 92918  | 43 AC020050 | AC020050 Drosophil  |
| C 8        | 39.6    | 2.7   | 123296 | 51 AC008323 | AC008323 Drosophil  |
| C 9        | 39.6    | 2.7   | 130583 | 41 AC007420 | AC007420 Drosophil  |
| C 10       | 39.6    | 2.7   | 304383 | 34 AE003579 | AE003579 Drosophil  |
| C 11       | 39.4    | 2.7   | 118593 | 10 HS288L1  | 282196 Human DNA s  |
| C 12       | 39.2    | 2.7   | 10619  | 2 AE002546  | AE002546 Neisseria  |
| C 13       | 39.2    | 2.7   | 133144 | 42 AC015424 | AC015424 Drosophil  |
| C 14       | 39.2    | 2.7   | 133190 | 51 AC010003 | AC010003 Drosophil  |
| C 15       | 39.2    | 2.7   | 154381 | 51 AC009369 | AC009369 Drosophil  |
| C 16       | 39.2    | 2.7   | 271237 | 34 AE003521 | AE003521 Drosophil  |
| C 17       | 37.6    | 2.6   | 503    | 5 I33795    | I33795 Sequence 15  |
| C 18       | 37.6    | 2.6   | 3900   | 5 I33788    | I33788 Sequence 6   |
| C 19       | 37.6    | 2.6   | 3902   | 33 DMU31226 | U31226 Drosophila   |
| C 20       | 36.8    | 2.6   | 29692  | 42 AC013145 | AC013145 Drosophil  |
| C 21       | 36.8    | 2.6   | 102619 | 33 DMBR37P7 | AL050331 Drosophil  |
| C 22       | 36.8    | 2.6   | 314661 | 34 AE003417 | AE003417 Drosophil  |
| C 23       | 36.6    | 2.5   | 39228  | 1 MLCB1788  | AL008609 Mycobacte  |
| C 24       | 36.6    | 2.5   | 167306 | 32 AL133507 | AL133507 Homo sapi  |





```
Db 1141 ATACACCGTGGCGGATGCTGAACAAACAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
Qy 1201 tgaacataccggatgtagacgcgggtaagtgtgaagtcaacaacaatcgctatttttaa 1260
Db 1201 TGAACATACCGGATGATAGATCGGGTAATGTGTAAAGTTCAACAACATCGCTATTTTAA 1260
Qy 1261 sagctaaagcagggtgcataatgagggccagatacaccccatcaatattggtttactactec 1320
Db 1261 CAGCTAAAGCAGGTGCATATGGGGCCAGATACACCCATCAATATTGGTTTACTTACTCC 1320
Qy 1321 ttcaagcagtagacggcgacaaagagttgcacaatggcgagcaacccagcattt 1380
Db 1321 TTCACGGAGTAGCGCGGACCAAGAGTTGTACAAATGGCGGGAGCAACCCAGGCTATT 1380
Qy 1381 gccgaataatacaaaatggcgcatcaacggcagaccactcaatgcaattcattggat 1440
Db 1381 GCCGAATTAATCAAAATGGCGCATCAACGGCAGACCACCTCAATGCAATTCATTGGAT 1440
Qy 1441 cc 1442
Db 1441 CC 1442

RESULT 2
A72152 LOCUS A72152 1442 bp DNA PAT 11-MAY-1999
DEFINITION Sequence 1 from Patent WO9801568.
ACCESSION A72152
VERSION A72152.1 GI:4808107
KEYWORDS
SOURCE Klebsiella oxytoca.
ORGANISM Klebsiella oxytoca
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Klebsiella.
REFERENCE 1 (bases 1 to 1442)
AUTHORS Brieden, W., Naughton, A., Robins, K., Shaw, N., Tinschert, A. and
Zimmermann, T.
TITLE METHOD OF PREPARING (S) - OR (R) -3,3,3-TRIFLUORO-2-HYDROXY-2-
METHYLPROPIONIC ACID
JOURNAL Patent: WO 9801568-A 15-JAN-1998;
LONZA AG (CH)
FEATURES
source
1..1442 Location/Qualifiers
/organism="Klebsiella oxytoca"
/strain="PRS1"
/isolate="PRS1"
/db_xref="taxon:571"
197..1183
/codon_start=1
/transl_table=11
/product="AMIDASE"
/db_xref="GI:4808108"
/translation="MKWLEESIMAKRGVAGRKPVTHLHTEMOKEPHYHTGYPSTPV
LTIEPGRDIIVDTRDAFEGA INSEODIPSQLKMPFLNPONGP INVNAEKGDLVAVY
TESMLPRGVGYGCAMTHPEGLTGDTLAMLNDPLPEKVRMIKIDSEKVKSRHT
LPYKPHIGTSLSPSEIDINSILTPDNHGNMDVPDIGPSITYLPVPAPGGRFLTGDA
HACODGRTICATVEFASITTIKVDLIKMWLSWRMENAINISIGSARPLEDATRI
AYRDLIWLVEDFGFEQWDAYMLLSQCKVRLGNMVDPKYTVGAMLKNLLV"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN

Query Match 100.0%; Score 1442; DB 5; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccgggaactccatgtggcgtgatcctgtgcagcagatattgcgatgccagcggg 60
Db 1 CCCGGNACTCATGTGGCGGTGATCCTGTGTCGACGAGATATTGCGATGATCCAGCGG 60
Qy 61 ccgcacagcgtgtggttaataaggcctggttagaagcgtgaaccaaac 120
Db 1141 ATACACCGTGGCGGATGCTGAACAAACAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
```

```
Db 61 CCGCACAGCGCTGTGCGGTAATGGATAAAAGGCGCTGGTTGTAGAAACCGCTGACCCAAACAAC 120
Qy 121 agctctctgatgatcttttaataatgcgtcatctgctctgtaactaaacgcgtataaatt 180
Db 121 AGCTCTCTGATGATCTTTTAATGGCGTGCATCTGGCTCTGTAACCTAAACGCTATAAATT 180
Qy 181 acgtggagaataacatatgaaatggttggaagaatccattatggccaaacggggtgttg 240
Db 181 ACGTGGAGAATAACATATGAAATGGTTGGAAGAAATCCATTATGCGCCAAACGCGGTGTGG 240
Qy 241 tcccgggcgtaaacggtaacgcatcacctgacggaagaataatgcaaaaagagtttcatia 300
Db 241 TGCCGGCGCTAAACCGGTAAACGCATCACCTGACGGGAAGAAATGCAAAAAGAGTTTCATTA 300
Qy 301 cccatttggcccttattccacacccqctctgaccatcgaacccggtgacgcgattattgt 360
Db 301 CACCATTGGCCCTTATTCCACACCCGCTCTGACCATCGAACCCTGGTGACCGGATATTGT 360
Qy 361 cgactcgagatgcttttgaagggtgctatcaattcggcaacaggatatattccgagccagtt 420
Db 361 CGACACTCGAGATGCTTTTGAAGGTGCTATCAATTCGAACAGAGATATTCCGAGCCAGTT 420
Qy 421 gctaaaaatgcccttctcaacccacacaaacggaccgcatcgtgctcaatggcgcgagaa 480
Db 421 GCTAAAAATGCCCTTTCTCAACCCACAAAACGACCGATCATGGTCAATGGCGCGGAGAA 480
Qy 481 agtgatgtgctgcgtgctctatcgaaatccatggttgcctcgcggtgatccctacgg 540
Db 481 AGGTGATGTGCTCGCTGCTCTATATCGAATCCATGTTGCCCGGGCGTGTGATCCCTACGG 540
Qy 541 catctgcgcattgctccgcatctttggcggaactgacgggacgcacctgacggccatgct 600
Db 541 CATCTGCGCATGATTCGCGATTTTGGCGACTGACGGGACCGAGCTGACGGCCATGCT 600
Qy 601 caatgctccgctgcagaaaagggtgcgcatgattaaactcgacagtgaaagctactg 660
Db 601 CAATGATCCGCTGCCAGAAAAGGTGCGCATGATTAACCTCGACAGTGAAAAGGTCFACTG 660
Qy 661 gagcaaacgccatcacgcttccctataaaccctattgacaccttggcgttagcgtatcgccaga 720
Db 661 GAGCAAAACGCCATACGCTTCCCTATAAACCCCATATTGGCACCTTGAGCGTATCGCCAGA 720
Qy 721 aattgactcaataattcactgacgcccagacaatcacgcggggaataatggatgctgcgga 780
Db 721 AATTGACTCAATCAATTCACTGACGCGCAGACAAATCACGGCGGAATATGGATGTGCGGA 780
Qy 781 tataggaccaggagattacacctatctgcggtacgtgcgctggagggcgccgtgttat 840
Db 781 TATAGGACCGAGGAGTATTACCTTATCTGCGGTACGTGGCCCTGGAGGGCGCGCTGTTAT 840
Qy 841 tggatgcccctgctgttcagggtgatggtgagatttgcgggacgcagtagagttgc 900
Db 841 TGGTGATGCCATGCTGTGTCAGGGTGATGGTGAGATTTGCGGGACCGCAGTAGAGTTGC 900
Qy 901 ctcaatcacacacatcaaaagtgcatttgcataagaactggcagcttctctggccacgaat 960
Db 901 CTCAATCACACCATCAAAAGTCGATTGTATCAAGAAGCTGGCAGCTTCTCTGGCCACGAAT 960
Qy 961 ggagaatgcggaaaaatattatgagattggcagtcgacgtccgctggagagatcgacgcg 1020
Db 961 GGAGAATGCCGAAAATATTATGAGTATTGGCAGTCCACGCTCCGCTGGAGATCGGACGG 1020
Qy 1021 aattgcataatcgagcttaattactgctggttagaagactttggtctggaacaatggga 1080
Db 1021 AATTGCATATCGGACTTTAATTTACTTGGCTGGTAGAGACTTTGGCTTCGAACAATGGGA 1080
Qy 1081 tgcctacatgcttctgagtcgaatgcggcaaatgcggctgggcaacatgctgcaccccaa 1140
Db 1081 TGCTACATGCTTCTGAGTCAATGCGCAAAAGTGGCGTGGGCAACATGGTCCGCCCCAA 1140
Qy 1141 atacacgctgtgcgcatgctgaacaaaaacctgttagtttagtagaataaataaccgg 1200
Db 1141 ATACACCGTGGCGGATGCTGAACAAACAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200
```

|  |      |  |      |      |
|--|------|--|------|------|
| QY   | 1201 | tgaacattaccggatgtagatgggtaatagtgtgaagttcaacaatacgtatttttaa | 1260 | gene |
|  |      |  |      |      |
| CDS  |      | complement(422..595)                                       |      |      |
| Db   | 1201 | TGAACATTACCGGATGATGTCGGGTAATGTGTAAGTTCAACAATCGTATTTTAA     | 1260 |      |
|  |      |  |      |      |
| QY   | 1261 | cagctaaagcaggtgcataatggggccagatcacacccatcaattgtttactttacc  | 1320 |      |
|  |      |  |      |      |
| Db   | 1261 | CAGCTAAAGCAGGTGCATATGGGGCCAGATACACCCATCAATATGGTTTACTTCC    | 1320 |      |
|  |      |  |      |      |
| QY   | 1321 | ttcagcgagtgacggcgccacagagttgtcacatggcgcgagcaacccagggctatt  | 1380 |      |
|  |      |  |      |      |
| Db   | 1321 | TTCAGCGGAGTGACGGCGGCACAGAGTTGTCAACATGGCGGAGCAACCCAGGCTATT  | 1380 |      |
|  |      |  |      |      |
| QY   | 1381 | gccgaataatcaaaatggcgccatcaacggcgagacaccatcaatgaattcatttgat | 1440 |      |
|  |      |  |      |      |
| Db   | 1381 | GCCGAATTAATCAAAATGGCGGCATCAACGGCAGACCACTCAATGCAATTCATTGGAT | 1440 |      |
|  |      |  |      |      |
| QY   | 1441 | cc 1442  |      |      |
|  |      |  |      |      |
| Db   | 1441 | CC 1442  |      |      |
| RESULT 3   |      |  |      |      |
| AP000059/c   |      |  |      |      |
| LOCUS  |      |  |      |      |
| DEFINITION   |      |  |      |      |
| ACCESSION  |      |  |      |      |
| AP000059.1 GI:5103911  |      |  |      |      |
| KEYWORDS   |      |  |      |      |
| SOURCE   |      |  |      |      |
| ORGANISM   |      |  |      |      |
| Aeropyrum pernix (strain:K1) DNA.                                  |      |  |      |      |
| Aeropyrum pernix   |      |  |      |      |
| Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;     |      |  |      |      |
| Aeropyrum.   |      |  |      |      |
| 1 (sites)  |      |  |      |      |
| Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Halkawa,Y.,    |      |  |      |      |
| Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H.,  |      |  |      |      |
| Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Nakazawa,H.,        |      |  |      |      |
| Takamiya,M., Masuda,S., Funahashi,T., Tanaka,T., Kudoh,Y.,         |      |  |      |      |
| Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K., Kubota,K.,            |      |  |      |      |
| Nakamura,Y., Nomura,N., Sako,Y. and Kikuchi,H.                     |      |  |      |      |
| Complete genome sequence of an aerobic hyper-thermophilic          |      |  |      |      |
| crenarchaeon, Aeropyrum pernix K1                                  |      |  |      |      |
| DNA Res. 6 (2), 83-101 (1999)                                      |      |  |      |      |
| 99310339   |      |  |      |      |
| 2 (bases 1 to 167000)  |      |  |      |      |
| Tanaka,T., Hino,Y., Kawarabayasi,Y. and Kikuchi,H.                 |      |  |      |      |
| Direct Submission  |      |  |      |      |
| Submitted (14-DEC-1998) to the DDBJ/EMBL/Genbank databases. Yutaka |      |  |      |      |
| Kawarabayasi, National Institute of Technology and Evaluation,     |      |  |      |      |
| Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo         |      |  |      |      |
| 151-0066, Japan (E-mail:kyutaka@kazusa.or.jp, Tel:+81-3-3481-8951, |      |  |      |      |
| Fax:+81-3-3481-8424)   |      |  |      |      |
| Kawarabayasi, Y. is officially affiliated with the National        |      |  |      |      |
| Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki     |      |  |      |      |
| 305-0046, Japan.   |      |  |      |      |
| 606-8502, Japan  |      |  |      |      |
| The other authors are at the National Institute of Technology and  |      |  |      |      |
| Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.     |      |  |      |      |
| All the sequence with length 100 codons or more between ATG or GTG |      |  |      |      |
| and stop codon are defined as CDS                                  |      |  |      |      |
| Homology analysis is performed by Smith-Waterman algorithm against |      |  |      |      |
| Genbank and Genpept release 109; EMBL release 56.0; Swissprot      |      |  |      |      |
| release 36.0; PIR-Protein release 57.0; and OWL release 31.0.      |      |  |      |      |
| E-mail address for comments and questions: genome@ap.nite.go.jp    |      |  |      |      |
| Restriction map, ORF organization, sequence alignment and more     |      |  |      |      |
| information are available at W.W.W. site of Biotechnology Center,  |      |  |      |      |
| URL: http://www.mild.nite.go.jp/.                                  |      |  |      |      |
| Location/Qualifiers  |      |  |      |      |
| 1..167000  |      |  |      |      |
| /organism="Aeropyrum pernix"                                       |      |  |      |      |
| /strain="K1"   |      |  |      |      |
| /db_xref="taxon:56636"   |      |  |      |      |
| FEATURES   |      |  |      |      |
| source   |      |  |      |      |
| complement(422..595)   |      |  |      |      |
| /gene="APES009"  |      |  |      |      |
| complement(422..595)   |      |  |      |      |
| /gene="APES009"  |      |  |      |      |
| /note="similar to PIR:H71068 percent identity:29.167 in            |      |  |      |      |
| 48aa."   |      |  |      |      |
| /codon_start=1   |      |  |      |      |
| /transl_table=11   |      |  |      |      |
| /product="57aa long hypothetical protein"                          |      |  |      |      |
| /protein_id="BAA79228.1"   |      |  |      |      |
| /db_xref="GI:5103912"  |      |  |      |      |
| /translation="METAIVYHLANKFHPAIIAILGVSDYKAKNKVEFVYEBELKDE          |      |  |      |      |
| IFEIGLELTLEVK"   |      |  |      |      |
| complement(1152..1490)   |      |  |      |      |
| /gene="APE0276"  |      |  |      |      |
| complement(1152..1490)   |      |  |      |      |
| /gene="APE0276"  |      |  |      |      |
| /note="similar to PIR:F69324 percent identity:47.115 in            |      |  |      |      |
| 104aa."  |      |  |      |      |
| /codon_start=1   |      |  |      |      |
| /transl_table=11   |      |  |      |      |
| /product="112aa long hypothetical protein"                         |      |  |      |      |
| /protein_id="BAA79229.1"   |      |  |      |      |
| /db_xref="GI:5103913"  |      |  |      |      |
| /translation="MRAGFIRVEEVDDRLVSLFEEFVDTGEGAEIVLAFERNAGLLL          |      |  |      |      |
| MDDRDARNLAKLGLQVMGTGLVIALAKYGLTSKAKPIIDKLIESGFWSIRRTLEEF           |      |  |      |      |
| LRKLGSRNK"   |      |  |      |      |
| 1228..1722   |      |  |      |      |
| /gene="APE0277"  |      |  |      |      |
| 1228..1722   |      |  |      |      |
| /gene="APE0277"  |      |  |      |      |
| /codon_start=1   |      |  |      |      |
| /transl_table=11   |      |  |      |      |
| /product="164aa long hypothetical protein"                         |      |  |      |      |
| /protein_id="BAA79230.1"   |      |  |      |      |
| /db_xref="GI:5103914"  |      |  |      |      |
| /translation="MSLSIIGFALEVNPLYLANIIPSVDPITCNPSFLARFLASRSI          |      |  |      |      |
| SSKPAIRSKARTIASPVSNTSKRDTLSLTSTSTRIKPALTTPSLGLPATVISA              |      |  |      |      |
| QTVSGTTRSPKFSRRSRRTLASAMIMLETTTRSLSSSSLLLLKLLGLQLQLLL              |      |  |      |      |
| LAIV"  |      |  |      |      |
| complement(2280..2756)   |      |  |      |      |
| /gene="APE0278"  |      |  |      |      |
| complement(2280..2756)   |      |  |      |      |
| /gene="APE0278"  |      |  |      |      |
| /codon_start=1   |      |  |      |      |
| /transl_table=11   |      |  |      |      |
| /product="158aa long hypothetical protein"                         |      |  |      |      |
| /protein_id="BAA79231.1"   |      |  |      |      |
| /db_xref="GI:5103915"  |      |  |      |      |
| /translation="MDKEKDGAEVLDTNIIILASVLRHGSYTRQWVYILIDVLGIK           |      |  |      |      |
| AFTSEKALHEIDEHELEAKRKVSVEELRAAIRILLNNVMEKGTYNAYMEKAGEC             |      |  |      |      |
| VKDLADAFALALHILSEKYKNVILLTWNKNKYIENCIRKAYGVIITYTPSDLRVPN"          |      |  |      |      |
| complement(3301..3669)   |      |  |      |      |
| /gene="APE0279"  |      |  |      |      |
| complement(3301..3669)   |      |  |      |      |
| /gene="APE0279"  |      |  |      |      |
| /note="similar to PIR:H69385 percent identity:39.175 in            |      |  |      |      |
| 97aa."   |      |  |      |      |
| /codon_start=1   |      |  |      |      |
| /transl_table=11   |      |  |      |      |
| /product="122aa long hypothetical protein"                         |      |  |      |      |
| /protein_id="BAA79232.1"   |      |  |      |      |
| /db_xref="GI:5103916"  |      |  |      |      |
| /translation="MVCIEPHSISVYAPRLFLVEVAGLVRYLAPLSIVERVLDAFSS          |      |  |      |      |
| KVLVDGEAFRIAEVALATAGSGADSYILGLAKTLNLPVATSDKVQAQNAKAGIK             |      |  |      |      |
| SFYILSNNELEELMKYMCCK"  |      |  |      |      |
| complement(3752..3979)   |      |  |      |      |
| /gene="APES010"  |      |  |      |      |
| complement(3752..3979)   |      |  |      |      |
| /gene="APES010"  |      |  |      |      |
| /note="similar to PIR:F69385 percent identity:38.462 in            |      |  |      |      |
| 52aa."   |      |  |      |      |
| /codon_start=1   |      |  |      |      |

```
/transl_table=11
/product="75aa long hypothetical protein"
/protein_id="BAA79233.1"
/db_xref="GI:5103917"
/translation="MEKRRVEIGMSKVIRVRYEKVLKPLEPVNLEDGEEVDIIIREN
LAELARRIRRLSOREEPSEILSRERLA"
4028..4402
/gene="APE0280"
CDS
4028..4402
/gene="APE0280"
/codon_start=1
/transl_table=11
/product="124aa long hypothetical protein"
/protein_id="BAA79234.1"
/db_xref="GI:5103918"
/translation="MIRSPDYHISWCVPFYNAHMOHSYLYSAFSGAAGPWGNSPRRAR
DKTYITINCLILNHLFPDGIHTCPKSQIEPPPEELLYRLDYLNMVLTLRHNLSGG
MYVALEGGLSIVGIDMEGDLR"
complement(4516..4842)
/gene="APE0281"
complement(4516..4842)
/gene="APE0281"
/codon_start=1
/transl_table=11
/product="108aa long hypothetical protein"
/protein_id="BAA79235.1"
/db_xref="GI:5103919"
/translation="MVTHTPTVTIVLSPMTTSSGVNRMWAGTVPAGNNATEHTAVPAISO
SSDFSTFCASRTSIAASSAVVSALPTPTPATILTSSTKLLYLGLYFSPPHDKIRKAN
KRWAP"
complement(4803..5465)
/gene="APE0282"
complement(4803..5465)
/gene="APE0282"
/codon_start=1
/transl_table=11
/product="220aa long hypothetical protein"
/protein_id="BAA79236.1"
/db_xref="GI:5103920"
/translation="MSSGILFOYRSNICVYFFHFGSSLFASLIASISMDIKGLFIW
SAYARVLPWVFLAGRACAIISFAYSOLAFRMWYILSLFTPCFPASFRVIAVSLQGG
LISLSFLNSWYSWFMCNLFMWGSSASFSLFRSRYALSYSAKSCSYASSIVSTMY
WSICSRVSWGALSSMCCFFFLFLRSRMSMPAPRPISLLILPSLWSPQIRUPCY
PR"
4831..6840
/gene="APE0283"
4831..6840
/gene="APE0283"
/feature="motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="669aa long hypothetical protein"
/protein_id="BAA79237.1"
/db_xref="GI:5103921"
/translation="MGDRHGDRMSRELIGAGILIRLLLNKKKKQOHIIEAPPQ
LTRQIDQYIVDTIDEAYEQAEYDRAYRELLKREKEADPHIKLIQNIKEYHEL
FRKNEMPPWMLTALLNEALKHGVKDRDKIYQVILKASWEAKETIAQALPAKYYQG
RYTRAYADQMRNPFTSIIETRAMIREANRDPDKYKQTMFDLYWNKIPDDIYNRADS
OHHEVTKYLLLEALREARESGVEPPWHKTPPEILIGPGECPAGEPLIAYPDR
YPTLQAKVLLTLVGEFLREGGVVHRDLYRALRNLLDSPLTNNHWPFLAEWFENLE
KGRHVSPTLQQLKLPVNPQLQRAYRYLAQIEQKGSIEIPEERKAFREVIEWLKN
EIVNRDFRVFVIEELVATGTMNKINDAALIGVLDLDFAKIYSGEKSPQAGIERDAT
QVLAVIKNLRDGNVDPYHVHTLINIEQHIQDPLLEAYKRTLEHYQTHGIPR
DLRYLAEERLEERLEKLGLENRNKLSGDIRRTAMLYRTMLLPHPSLIEDIKR
FYSYLAERLKVYGEESLPVHGVPAEDVASSFAIRAKPEERVRVRRREISGFEPL
FHAPHIKGVMSVLAATRSVSGEKHEEFIRHLERTQVRRLPE"
complement(5360..5668)
/gene="APE0284"
complement(5360..5668)
/gene="APE0284"
/codon_start=1
/transl_table=11
/product="102aa long hypothetical protein"
```

```
/protein_id="BAA79238.1"
/db_xref="GI:5103922"
/translation="MSGTCPPRRVSALPPARDPLVLCRGFPRGSCARRALPPIPLFP
HEGPLPGVAGIILLHDAASRPYRCRPECSCNTDQTSVYFSTSDHPCLLLSSPRFL"
Query Match 3.2%; Score 45.8; DB 1; Length 167000;
Best Local Similarity 53.0%; Pred. No. 0.07;
Matches 98; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 756 atgcggcggaatggtgctgcggatataggaccaggagattacctatctcccggtac 815
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 161824 ACGAGGAAACCTCGATGTGAGGCATCTCCCCAGGCTCCAAAATCTACTCCCGTGT 161765
QY 816 gtgcgcctgagagccgcctgtttattggtgatcccatgcttgcaggggtgatggtaga 875
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 161764 TCGTCGAGGGGGCGCTATTCTCGGTGGCGCACGCCACTATGCCAGGGCGGCGGAGG 161705
QY 876 ttgcgggaccgcagtagagttgtcctcaatcacccaccatcaaaagtcgattgatcaaga 935
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 161704 TGTGCGGCACAGCCATAGAGATGGGGCCACCGCTACCTTAAGGTTTCGGGGTCATAAGCG 161645
QY 936 actgg 940
Db 161644 AGGGG 161640
RESULT 4
SPAC869 35346 bp DNA PLN 15-DEC-1999
LOCUS S.pombe chromosome 1 cosmid c869.
DEFINITION AL132779
ACCESSION AL132779.2 GI:6594223
VERSION alpha-galactosidase; amidase; amino-acid permease;
KEYWORDS flavohemoprotein; formamidase; globin; oxidoreductase;
protein-l-isopartate o-methyltransferase; sodium:solute symporter
family; sulfate permease; sulfate transporter family; TFI LTR; urea
active transporter.
fission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
Schizosaccharomycetaceae; Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 35346)
AUTHORS Hunt,C., Aves,S., McDougall,R.C., Rajandream,M.A. and Barrell,B.G.
TITLE Submitted (02-NOV-1999) European Schizosaccharomycetes genome
JOURNAL sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and Department of Biological Sciences, University of Exeter, Perry
Road, Exeter EX4 4QG, United Kingdom
On Dec 16, 1999 this sequence version replaced gi:6224591.
COMMENT
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was
sequenced by the Sanger Centre. The sequencing of the S. pombe
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
```

Tue Sep 12 08:47:29 2000

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

# FEATURES

```

source
    location/Qualifiers
    1. .35346
        /organism="Schizosaccharomyces pombe"
        /strain="972h-"
        /db_xref="taxon:4896"
        /chromosome="I"
        /clone="cosmid c869"
        /map="IR"
        complement(1. .1610)
        /gene="SPAC869.11"
    1. .1141
        /note="Overlap with c922 S. pombe chromosome 1"
        complement(1. .1610)
        /partial
        /gene="SPAC869.11"
        /note="SPAC869.11, len:536, SIMILARITY:Schizosaccharomyces
        pombe, ISP5_SCHPO, sexual differentiation process putative
        amino-acid permease isp5., (543 aa), fasta scores: opt:
        1812, E():0, (53.9% identity in 492 aa)"
        /codon_start=1
        /label="SPAC869.11"
        /product="amino-acid permease"
        /protein_id="CAB60021.1"
        /db_xref="GI:6224602"
        /translation="MEPEYVSFSDKTSILNESKSLKDVKPSLEKSVITPGLVDD
        VEPKGNVRFDDPAKATGDTALKRSLSRHMOMISIGGATGTYVGSSS
        LADGSPASVILNYSLIGIMPEFIYALGEMAVYPVAGGENTVATRIDPAGFAVSW
        NYFINVFPELLEITCAIFRYWTDINSAAWISILVLIIVNLCVRAVGEVFIL
        STKVATFPIILAILINGGVPTDHRGYIGGSLIKHPRHGKFGCSEVTAARS
        FSGTEVIGLAAEVDNPKALPHAVQVFWRIALFYVVSLLIGLLSPDDPNLMNGS
        STVSPEFVLAIEKANIKGLPSVFNAAVLIISVYTSNSTYTAGRTLHGMANLKQPSF
        FKYIDRLGRILLAMVILLFFGFAYINEADKNGNDVSTVFNNLLALSGLSNFTWGS
        ICLGHIIPRIAPKQGHSLKELGFVSPMGIMGSCIGLFFNLICLMAQFYVSLFPIGK
        PNANDFGVGLAAPVTLAFTFYGIKYIDRS"
        complement(3. .1394)
        /gene="SPAC869.11"
        /note="Match to PF00324 aa_permeases, Amino acid permease
        Score 820.26"
        /gene="SPAC869.10c"
        3705. .5363
        /gene="SPAC869.10c"
        3705. .5363
        /gene="SPAC869.10c"
        /note="SPAC869.10c, len:552, SIMILARITY:Emericella
        nidulans, PUTX_EMENI, proline-specific permease, (550 aa),
        fasta scores: opt: 1968, E():0, (53.3% identity in 520
        aa)"
        /codon_start=1
        /label="SPAC869.10c"
        /product="amino-acid permease"
        /protein_id="CAB60020.1"
        /db_xref="GI:6224601"
        /translation="MDEKALQYEEESKRAEITDIELVSVGIDVEKKYGETKRALKS
        RVOLTAIGCICTGLFVGSGSALSESQPSALFSLVIMSFVITVMALGEMCTYLP
        LSGASPTIERYVDASLAFAAGWNWYAVFLVASEVTAASIVIEYWTYAVPTGATY
        ALLFLVAVLNSFFVKWGETEFPAIKVIAIGLILGVVIFFGTPKHDLRGFRY
        WKHGLAFRETYIKFSGSRVGFWSAVIKSGFAFILAPELVIFSAGETEAPRNIPKAT
        SRFIVRLIFYIFGSLITGVITSSKDDPRLNLAISSGASGAASAPVIGIQNAEIPVLN
        HTINAVILTSACSSNSFLFAGRSIYSLAKEHQAPKIFKYNCRNGVPVIVSVAIVLP
        ACIALNASAAVNFNWFNCNLSITISGLAWICVAVLOQFRAKILANNMETRYPKT
        PPQPYATLITLLALITNGFTVFGHTFAGNFIAAYITPLIFLVLYVAHLKWSR
        NRSFGKRIEIDVTIGVAEAEALQMPYAPVPNRNIEKIWFIA"
        3816. .5228
        /gene="SPAC869.10c"
        /note="Match to PF00324 aa_permeases, Amino acid permease
        Score 718.22"
        LTR
        complement(6788. .7138)
        /note="TFL LTR"
        gene
        complement(7524. .7874)
        /gene="SPAC869.09"
    
```

CDS

```

complement(7524. .7874)
/gene="SPAC869.09"
/note="SPAC869.09, len:116, SIMILARITY:Schizosaccharomyces
pombe, YAAB_SCHPO, hypothetical 15.4 kd protein c22g7.11c
in chromosome I., (140 aa), fasta scores: opt: 192,
E():2.5e-06, (39.6% identity in 139 aa)"
/codon_start=1
/label="SPAC869.09"
/product="hypothetical protein"
/protein_id="CAB60019.1"
/db_xref="GI:6224600"
/translation="MPDPAHIIAGHKAALSNPHVSEAEKERARKYLKHGSHSHYTTG
TTRGKADADDAGELREEGFTKNQFEDNESQAKNLGNVRGYKAAMHNPKVSKGRR
HAKELLEVDDESK"
complement(8756. .9448)
/gene="SPAC869.08"
complement(8756. .9448)
/note="SPAC869.08, len:230, SIMILARITY:Triticum aestivum,
PMT_WHEAT, protein-l-isoaspartate o-methyltransferase,
(230 aa), fasta scores: opt: 602, E():2.2e-32, (48.5%
identity in 200 aa)"
/codon_start=1
/label="SPAC869.08"
/product="putative protein-l-isoaspartate
o-methyltransferase"
/protein_id="CAB60018.1"
/db_xref="GI:6224599"
/translation="MEFNSFLSSNAALVQHVLESKFLTNORAIKAMNATSRGFCPLS
PYMDSQSGYGYTISAPHHMATAQLEFVLQPGCSALDIGSGYLVAAWAMVAP
NHTVKGIEHPQLVETSKNKLADINHDEVLMEMYKEKRLQINGVGRMGTSDEKFD
AIHVGSASSELKQKLVLDQKPGKILIPITIGTYSQNIYLIEKNEQGISKRTLPFPRYV
PLTDSDDSDY"
complement(8783. .9427)
/gene="SPAC869.08"
/note="Match to PF01135 PCMT,
protein-L-isoaspartate(D-aspartate) O-methyltransferase
(PCMT) Score 263.14"
10265. .11575
/gene="SPAC869.07c"
10265. .11575
/gene="SPAC869.07c"
/note="SPAC869.07c, len:436, SIMILARITY:Mortierella
vinacea., Q02402, alpha-galactosidase, (417 aa), fasta
scores: opt: 1443, E():0, (52.3% identity in 407 aa)"
/codon_start=1
/label="SPAC869.07c"
/product="putative alpha-galactosidase"
/protein_id="CAB60017.1"
/db_xref="GI:6224598"
/translation="MISIFLNCFFLVFLFTSDVHGSYNGLGKLPQMGWSNWKYA
CDDIESIILNNAKAKEEGLLDGYEIVVMDDCWSKHENATTTGRLEANPKPPNGIG
SMARKLHDMGFKFGYSSAGKYTCAGFPGLSNGHEQIDADTFADGVDYLYKDCNFEG
KSGVPLISYERYKRMSDALNKTGRPIFYLCQMGEDFVWNGNTIANSWRIISGDIPT
FSRKVDPCPETIECFALQGDHCSVMNIIISKASPLSSKAGMNSGNDLDSLEVGNGM
KRYEYKHTFMTAILKSPILGNDVSSMSPKLIVSNKLIISINQIDIGNPALIWK
KKYDEYIEFSGSLNNDVAVLNAASPELKMGIHLSDIFVDALGNAEHDWLATDL
WNNNVLVSDRIRANASHGVQVWFQYKVKNTDKFFSEKHI"
13058. .13669
/gene="SPAC869.06c"
13058. .13669
/gene="SPAC869.06c"
/note="SPAC869.06c, len:203"
/codon_start=1
/label="SPAC869.06c"
/product="hypothetical protein"
/protein_id="CAB60016.1"
/db_xref="GI:6224597"
/translation="MLOYSKKKVSILNFPFVRLLSVKMTFRISDAIFKHRLQSDYQNI
KSANDYDTRWQNFVALDHSVGEIVVYPKFVEGEGEMAEKDRHEQLVK
EMLYKFSQMKANQSNFIPALDELMSLQKHIDEEOHDIPELEKHLSEESLHASSF
ERTKKFVPTHSPSNKPPFETVAGLFAAIDIKLRDMMEKWP"
16383. .18905

```

misc\_feature

gene

CDS

gene

CDS

gene



us-09-214-679-1.rge

|                       |   |   |
|-----------------------|---|---|
| CDS                   |   | 10500.._11060<br>/note="PAR0445"<br>/codon_start=1<br>/transl_table=11<br>/product="hypothetical protein"<br>/protein_id="CA94569.1"<br>/db_xref="GI:5458080"   |
|                       |   | /translation="MSKFRVYHGDITRFKAELIYNAAKNKYLEHGSGGVAYAIKAASG<br>DVSEYIRISKEEMRKQIGDWIEHFVVTPPLNAKNGVKYVIHTVPYCCKGWDD<br>KRKLDAILALGAUKADELGLVRSIAFPALSIAGIYCPLEEIVKVTKLVNVEFLSKAKS<br>VTDYVLVYSERYDEVALKVLRDEL" |
| Query Match           | 2.9%; Score 41.2; DB 1; Length 307120;  |   |
| Best Local Similarity | 53.8%; Pred. No. 1.1.8;   |   |
| Matches               | Conservative 0; Mismatches 73; Indels 0; Gaps 0;  |   |
| QY                    | 743 acgccagacaatcacgcggcggaatatgagtgtcccgataagaaccaggagattacc 802   |   |
| Dn                    | 282090 ACCCCAGAAAAACATCGGGAAAACATGCACAAACCATACTAGAAGGAGCACCAT 282149  |   |
| QY                    | 803 tatctcgcgtagtcgccttggagcgcccttttatgtgatgcccatgcttgtcag 862  |   |
| Dn                    | 282150 TACTTCCCCTGTCTTGTTGATGGAGCTTACCTTCCTATAGCGATTGCCATGCCGTAATG 282209   |   |
| QY                    | 863 ggttgattgtagatttcgggacgcgcgtagatttgc 900  |   |
| Dn                    | 282210 GGAGTAGGGGAAGTCTCGGTCTCGGCTTGTGAGGTTTC 282247  |   |
| RESULT                | 6   |   |
| NMA2Z2491/c           | NMA2Z2491 349061 bp DNA BCT 30-MAR-2000   |   |
| LOCUS                 | Neisseria meningitidis serogroup A strain Z2491 complete genome;<br>segment 2/7.  |   |
| DEFINITION            |   |   |
| ACCESSION             | AL162753 AL157959   |   |
| VERSION               | AL162753.2 GI:7379120   |   |
| KEYWORDS              |   |   |
| SOURCE                | Neisseria meningitidis.<br>Neisseria meningitidis<br>Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;<br>Neisseria.   |   |
| REFERENCE             | 1 (bases 1 to 349061)<br>Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,<br>Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,<br>Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,<br>Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,<br>Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,<br>Skellton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.<br>Complete DNA sequence of a serogroup A strain of Neisseria<br>meningitidis Z2491 |   |
| JOURNAL               | Nature 404 (6777), 502-506 (2000)   |   |
| MEDLINE               | 20222556  |   |
| REFERENCE             | 2 (bases 1 to 349061)<br>Parkhill,J.  |   |
| AUTHORS               | Direct Submission   |   |
| TITLE                 | Submitted (30-MAR-2000) Submitted on behalf of the Neisseria<br>sequencing team, Sanger Centre, Wellcome Trust Genome Campus,<br>Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  |   |
| JOURNAL               |   |   |
| COMMENT               | Notes:<br>Details of N. meningitidis sequencing at the Sanger Centre are<br>available on the World Wide Web.<br>(URL, http://www.sanger.ac.uk/projects/N_meningitidis/).<br>Location/Qualifiers<br>1. 349061<br>/organism="Neisseria meningitidis"<br>/strain="Z2491"<br>/db_xref="taxon:487"<br>/note="serogroup: A"<br>complement(24..206)<br>/note="ATR repeat; hmms hit to HMM ATR (1 - 183), score:<br>310.39"<br>/label=ATR<br>209..212   |   |
| FEATURES              | source  |   |
|                       | repeat_unit   |   |
| RBS                   |   |   |

```

gene
220..681
/ gene="NMA0368"
CDS
220..681
/ gene="NMA0368"
/ note="NMA0368, probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/translation="MQEQRKSSPIYMLIVSVALWIASLNVAFYLGNGHSMGLTV
LILGSIASLDIRCYAVYANVWLAIVLLALRKVVPVHAAFGLGALVAFSVKAVYV
DEAGNTSDIVYAGFYLMWAAFAVASIGTTFAGNKKERKAASADGTNDV"
671..1492
/ gene="hemK"
CDS
671..1492
/ gene="hemK"
/ note="NMA0369, hemK, HemK protein, len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="HemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/translation="MTDFKWLGLSKLPKNARMLOVYSEYTRVOLLTRGGEMPDEV
RORADLAORLNGEPVAYILGAREFYGRRTVNPVSLIPRPEHLEAVLARLPEN
GRVWDLTGCSNAVTVALERPDFAFVRASDISPPALETARKNAADLGRVAFYGSWF
DTPMSEGRDIIIVSNPIYIENGDKHLSQGDLPFEPQIALTDFDGLSCIRTLAOGAP
DRLAEGGFLLEHGFDDQGAAVRGVLAENGFSGVETPLDLAGLDRVTLGKYMHLK"
831..840
/ gene="hemK"
misc_feature
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1162..1171
/ gene="hemK"
misc_feature
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1187..1207
/ gene="hemK"
misc_feature
/ note="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1565..1568
/ gene="NMA0370"
CDS
1578..2966
/ gene="NMA0370"
/ note="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P44640 hypothetical protein HI0325 (450 aa), fasta scores;
E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/translation="MNAVVVAVIVMLVLSLRVHVVLSTLTIGAFVGGAVAGMPLQNI
AAGQVSQAGIIPFNKLEGGAIALSYAMLGAFAMATHISGLPQOLAGAVVRKLN
RGMPDSVRSVGCVTKWLLSLITLVGMMSQNIPIPIAFIPMVPPLLLVNRKIDR
RLIACVITFGLVTVTMFLPYGFAIFLNEILLGNTHSAAPOLDVKNINVMMAIPL
GMLAGLLIAFVHYRKPRLYIQSNADTAGNADANRPQPSAYRSAAVAIAVCFAIOL
MYEISLVGLMGLGFVFMWLVINRKANDVFGIGIKMAMWGVFMIAAQGFPAVWNA
TGHIOPLVSSMAIFGNSKGMALAMLVVGLITMGISFSSTFLPIIAIYVPLCVGL
GFSPLATVAIVGTALGADGASPADSDTLGPTMGLNADQGDHIDRSVITPTIHNIP
LLIAGWIAAMVL"
2967..3191
/ gene="slyX"
CDS
2967..3191
/ gene="slyX"
/ note="NMA0371, slyX, SLX protein homolog, len: 74 aa;
similar to SLX_HAEIN P44759 SLX protein homolog (73 aa),
fasta scores; E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLX_ECOLI P30857 SLX protein (72 aa), fasta
scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="SLX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/translation="MDAVQEFHRTTELEIQSALQEDVIAGLNAMVAELRQTLDQQQ
QLRLLYQKMDRNPDAQEPYSLRDEIPHY"
complement(3271..3423)
/ gene="NMA0372"
CDS
complement(3271..3423)
/ gene="NMA0372"
/ note="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/translation="MFGRLSPGSDGFFLAVYEPVGVFICLOGKTAYFRSGGKRFRHRI
KGILSD"
3562..3571
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3586..4356)
/ gene="thiF"
CDS
complement(3586..4356)
/ gene="thiF"
/ note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
protein. (249 aa), fasta scores; E(): 0, 43.9% identity in
244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
pfam match to entry PF00899 ThiF_family, ThiF family"
/codon_start=1
/transl_table=11
/product="ThiF protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/translation="MTTTHDNDDAFLRYSRHLLDEIGEQOKLSAAHTLVVCGG
GLGAAALPYLAASGIGTLTIADSDTVLHNLQOVAFDEGDVGKLTALADRLRHIN
HTVDVRTINEKLDGCRLTGLVQADIVLCCDNATPQAVNRACVQAKTPLYSGAAVR
FEGQLAVYRPDLDPSPCYACLFDDGSGSDGICSLFGVFSPLVGIIGSTQAAEALKILL
DAGEPSSHGRLAYVRALEGWQYFDLPNRPCEPCVGAER"
3787..3796
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3847..4257)
/ gene="thiF"
misc_feature
/ note="Pfam match to entry PF00899 ThiF_family, ThiF
family, score 186.60, E-value 4.1e-52"
complement(4418..4427)
/ note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4471..7224
/ gene="ppc"
CDS
4471..7224
/ gene="ppc"
/ EC_number="4.1.1.31"
/ note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
917 aa; similar to many e.g. CAPP_RHOP A032483
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
Contains 2x Pfam match to entry PF00311 ppcase,
Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
site motif A (P-loop), and PS00393 Phosphoenolpyruvate
carboxylase active site 2"

```

```
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
/db_xref="GI:7379127"
/translation="NRHLRYIYSIYGNRKIMOLHILNPNKDAALAADEFLKQSLFN
LHHEASPLVETVYKLLSTSDSAALEIKVLPOLDEQTHDTLACGLFAQLINIAED
VHERRRQLHEAGRGGAESLTETVRRLKAGKADGKSVOROLDNTSVAVLTARHTE
VORQVLNFRIRALLPQRECTNADALRLREIDTILGLMOTSTRRLKLSVND
EINNGVSIFPKASFALPKYKMEHDFATYPOVVRVPOILKIGWISGDRDGNPFVS
AETLRFARRIADAVREFYRGELDKLYRELPLSIRRVKVGDMVALSKSPDETARA
EEPYRRATAYIMARANGKARALGLMGCKFGFELEPYASAQEFDLDLKLQLSLDNGS

Query Match      2.9%; Score 41.2; DB 2; Length 349061;
Best Local Similarity 53.0%; Pred. No. 1.8;
Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 188 gaataacatataaagtggtgaagaatccattatggaacacggtgtgtgtgcccggg 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106727 GCATCTTAACAAAACACGCAAAAGCCATGATGACCGAAGCGAGTTTATCCGCGCG 106668

QY 248 cgtaaacgggtacgcacatccatcgacggaagaatgcaaaaagagtttcattacaccatt 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106667 AGCGAAGCATTTTGAACACATCGAAGACCAATCGACGAAACGCGTGGGATTTCGAC 106608

QY 308 ggccttattcacacccgtcgtgaccatcgacacggtgacccggtgaccgga 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106607 TGCCGGTTGCCGGAACGTCCTGACCATCGAAGCGGAGCGGCA 106562

RESULT 7
AC020050
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION
VERSION
KEYWORDS
SOURCE
fruit fly
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 92918)
AUTHORS
Adams,M. and Venter,J.C.
TITLE
Direct Submission
JOURNAL
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10211861 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..92918
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 25719 a 20547 c 26010 g 26042 t
ORIGIN

Query Match      2.7%; Score 39.6; DB 43; Length 92918;
Best Local Similarity 61.8%; Pred. No. 4.2;
Matches 63; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 926 ttgatcaagaactggcagctttcttcgccaagatgagacgcaaaattattagtt 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49999 TTGCCCATGATCTGGGGTCCCATTTAGCCAAAAGGTAGCGCAAAATGCTCTCAGC 50058

QY 986 attgcagtcacgtccgctggagatgcgacgcgaattcca 1027
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 50059 ATCGGGGCATCAAGTTCCTGTCAGGATGCGCGGAGGGGAGCA 50100
```

RESULT 8  
AC008323  
LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 11, 2000 this sequence version replaced gi:5670417.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet

the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 92 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 840: contig of 840 bp in length

\* 841 920: gap of unknown length

\* 921 2166: contig of 1246 bp in length

\* 2167 2246: gap of unknown length

\* 2247 2984: contig of 738 bp in length

\* 2985 3064: gap of unknown length

\* 3065 3761: contig of 697 bp in length

\* 3762 3841: gap of unknown length

\* 3842 4489: contig of 648 bp in length

\* 4490 4569: gap of unknown length

\* 4570 5143: contig of 574 bp in length

\* 5144 5223: gap of unknown length

\* 5224 5878: contig of 655 bp in length

\* 5879 5958: gap of unknown length

\* 5959 6896: contig of 938 bp in length

\* 6897 8006: gap of unknown length

\* 8006 8086: contig of 1030 bp in length

\* 8087 9208: contig of 1122 bp in length

\* 8087

AC008323 132396 bp DNA HTG 10-FEB-2000  
Drosophila melanogaster chromosome 2 clone BACR03007 (D850) RPCI-98  
03.G.7 map 24D-24D strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 92 unordered pieces.

AC008323 2 GI:5957898  
HTG; HTGS\_PHASE1.  
fruit fly.

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 123296)  
Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,  
Kearney,L., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,  
Richards,S., Sethi,H., Swirskas,R.R., Wan,K.H., Webster,D.,  
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.

Sequencing of Drosophila melanogaster  
Unpublished  
2 (bases 1 to 123296)  
Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shnr,E.,  
Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.

Direct Submission  
Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 11, 2000 this sequence version replaced gi:5670417.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet

the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 92 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 840: contig of 840 bp in length

\* 841 920: gap of unknown length

\* 921 2166: contig of 1246 bp in length

\* 2167 2246: gap of unknown length

\* 2247 2984: contig of 738 bp in length

\* 2985 3064: gap of unknown length

\* 3065 3761: contig of 697 bp in length

\* 3762 3841: gap of unknown length

\* 3842 4489: contig of 648 bp in length

\* 4490 4569: gap of unknown length

\* 4570 5143: contig of 574 bp in length

\* 5144 5223: gap of unknown length

\* 5224 5878: contig of 655 bp in length

\* 5879 5958: gap of unknown length

\* 5959 6896: contig of 938 bp in length

\* 6897 8006: gap of unknown length

\* 8006 8086: contig of 1030 bp in length

\* 8087 9208: contig of 1122 bp in length

\* 8087



|          | Query Match   | 2.7%            | Score 39.6;    | DB 51;    | Length 123296; |
|----------|---|-----------------|----------------|-----------|----------------|
|          | Best Local Similarity   | 61.8%;          | Pred. No. 4.4; |           |                |
|          | Matches 63;   | Conservative 0; | Mismatches 39; | Indels 0; | Gaps 0;        |
| Qy 926   | ttgatcaagaactggcgagctttcttgcgcgcgaatggagaatgcgcgaataattatgagt | 985             |                |           |                |
|          |   |                 |                |           |                |
| Db 86793 | TTGGCCATGATCTGGGGTCCCATAGCCACAAAGGTGTAGGCCGAAATGCTCTCAGC      | 86852           |                |           |                |
| Qy 986   | attggcagtcgacgtccgctggagatgcgcgcgaattgca                      | 1027            |                |           |                |
|          |   |                 |                |           |                |
| Db 86853 | ATCGGGGATCAAGTTCGCTCAGGATCGCGGGAGGAGCA                        | 86894           |                |           |                |

## AC007420

LOCUS AC007420 130583 bp DNA HTG 20-SEP-1999  
 DEFINITION Drosophila melanogaster chromosome 2 clone BACR07M10 (D630) RPCI-98  
 07.M.10 map 24A-24D strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 83 unordered pieces.

## ACCESSION

VERSION AC007420.3 GI:5670587

## KEYWORDS

SOURCE HTG; HTGS\_PHASE1.

## ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 130583)  
 Celnikier,S.E., Agbavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chev,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.

## TITLE

Sequencing of Drosophila melanogaster

## JOURNAL

## REFERENCE

2 (bases 1 to 130583)  
 Celnikier,S.E., Agbavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chev,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.

## TITLE

Direct Submission

Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Aug 2, 1999 this sequence version replaced gi:5649319.

For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu). All contigs in this submission meet  
 the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 83 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 782: contig of 782 bp in length  
 \* 783 862: gap of unknown length  
 \* 863 1444: contig of 582 bp in length  
 \* 1445 1524: gap of unknown length  
 \* 1525 2743: contig of 1219 bp in length  
 \* 2744 2823: gap of unknown length  
 \* 2824 3716: contig of 893 bp in length  
 \* 3717 3796: gap of unknown length  
 \* 3797 4604: contig of 808 bp in length  
 \* 4605 4684: gap of unknown length  
 \* 4685 5390: contig of 706 bp in length  
 \* 5391 5470: gap of unknown length  
 \* 5471 6048: contig of 578 bp in length  
 \* 6049 6128: gap of unknown length  
 \* 6129 7034: contig of 906 bp in length  
 \* 7035 7114: gap of unknown length  
 \* 7115 7794: contig of 680 bp in length  
 \* 7795 7874: gap of unknown length  
 \* 7875 8799: contig of 925 bp in length  
 \* 8800 8879: gap of unknown length  
 \* 8880 9777: contig of 898 bp in length

9778 9857: gap of unknown length  
 9779 11049: contig of 1192 bp in length  
 9780 11129: gap of unknown length  
 9781 12047: contig of 918 bp in length  
 9782 12127: gap of unknown length  
 9783 12931: contig of 804 bp in length  
 9784 13011: gap of unknown length  
 9785 14016: contig of 1005 bp in length  
 9786 14096: gap of unknown length  
 9787 15426: contig of 1330 bp in length  
 9788 15507: gap of unknown length  
 9789 16623: contig of 1117 bp in length  
 9790 16703: gap of unknown length  
 9791 17527: contig of 824 bp in length  
 9792 17607: gap of unknown length  
 9793 18878: contig of 1271 bp in length  
 9794 18958: gap of unknown length  
 9795 19887: contig of 929 bp in length  
 9796 19967: gap of unknown length  
 9797 21188: contig of 1221 bp in length  
 9798 21268: gap of unknown length  
 9799 22540: contig of 1272 bp in length  
 9800 22620: gap of unknown length  
 9801 23401: contig of 781 bp in length  
 9802 23481: gap of unknown length  
 9803 24740: contig of 1259 bp in length  
 9804 24741: gap of unknown length  
 9805 25835: contig of 1015 bp in length  
 9806 25915: gap of unknown length  
 9807 27422: contig of 1507 bp in length  
 9808 27502: gap of unknown length  
 9809 28793: contig of 1291 bp in length  
 9810 28873: gap of unknown length  
 9811 30090: contig of 1217 bp in length  
 9812 30170: gap of unknown length  
 9813 31566: contig of 1396 bp in length  
 9814 31646: gap of unknown length  
 9815 33826: contig of 2180 bp in length  
 9816 33906: gap of unknown length  
 9817 35660: contig of 1754 bp in length  
 9818 3740: gap of unknown length  
 9819 36675: contig of 935 bp in length  
 9820 36755: gap of unknown length  
 9821 38411: contig of 1656 bp in length  
 9822 38491: gap of unknown length  
 9823 40297: contig of 1806 bp in length  
 9824 40377: gap of unknown length  
 9825 41875: contig of 1498 bp in length  
 9826 41955: gap of unknown length  
 9827 43789: contig of 1834 bp in length  
 9828 43869: gap of unknown length  
 9829 45471: contig of 1602 bp in length  
 9830 45551: gap of unknown length  
 9831 47128: contig of 1577 bp in length  
 9832 47209: gap of unknown length  
 9833 48263: contig of 1055 bp in length  
 9834 48343: gap of unknown length  
 9835 50567: contig of 2224 bp in length  
 9836 50647: gap of unknown length  
 9837 53132: contig of 2485 bp in length  
 9838 53212: gap of unknown length  
 9839 54731: contig of 1519 bp in length  
 9840 54811: gap of unknown length  
 9841 56680: contig of 1869 bp in length  
 9842 56760: gap of unknown length  
 9843 58314: contig of 1554 bp in length  
 9844 58394: gap of unknown length  
 9845 60268: contig of 1874 bp in length  
 9846 60348: gap of unknown length  
 9847 62780: contig of 2432 bp in length  
 9848 62860: gap of unknown length  
 9849 64898: contig of 2038 bp in length  
 9850 64978: gap of unknown length

|   |        |         |                              |
|---|--------|---------|------------------------------|
| * | 64979  | 66963:  | contig of 1985 bp in length  |
| * | 66964  | 67043:  | gap of unknown length        |
| * | 67044  | 69574:  | contig of 2531 bp in length  |
| * | 69575  | 69654:  | gap of unknown length        |
| * | 69655  | 72365:  | contig of 2711 bp in length  |
| * | 72366  | 72445:  | gap of unknown length        |
| * | 72446  | 75169:  | contig of 2724 bp in length  |
| * | 75170  | 75249:  | gap of unknown length        |
| * | 75250  | 78381:  | contig of 3132 bp in length  |
| * | 78382  | 78461:  | gap of unknown length        |
| * | 78462  | 81086:  | contig of 2625 bp in length  |
| * | 81087  | 81166:  | gap of unknown length        |
| * | 81167  | 83892:  | contig of 2726 bp in length  |
| * | 83893  | 83972:  | gap of unknown length        |
| * | 83973  | 86588:  | contig of 2616 bp in length  |
| * | 86589  | 86668:  | gap of unknown length        |
| * | 86669  | 91696:  | contig of 5028 bp in length  |
| * | 91697  | 91776:  | gap of unknown length        |
| * | 91777  | 94463:  | contig of 2687 bp in length  |
| * | 94464  | 94543:  | gap of unknown length        |
| * | 94544  | 98576:  | contig of 4033 bp in length  |
| * | 98577  | 98656:  | gap of unknown length        |
| * | 98657  | 101629: | contig of 2973 bp in length  |
| * | 101630 | 101709: | gap of unknown length        |
| * | 101710 | 112476: | contig of 10767 bp in length |
| * | 112477 | 112556: | gap of unknown length        |
| * | 112557 | 113457: | contig of 901 bp in length   |
| * | 113458 | 113537: | gap of unknown length        |
| * | 113538 | 114183: | contig of 646 bp in length   |
| * | 114184 | 114263: | gap of unknown length        |
| * | 114264 | 114990: | contig of 727 bp in length   |
| * | 114991 | 115070: | gap of unknown length        |
| * | 115071 | 115790: | contig of 720 bp in length   |
| * | 115791 | 115870: | gap of unknown length        |
| * | 115871 | 116529: | contig of 659 bp in length   |
| * | 116530 | 116609: | gap of unknown length        |
| * | 116610 | 117497: | contig of 888 bp in length   |
| * | 117498 | 117577: | gap of unknown length        |
| * | 117578 | 118139: | contig of 562 bp in length   |
| * | 118140 | 118219: | gap of unknown length        |
| * | 118220 | 118669: | contig of 650 bp in length   |
| * | 118670 | 118949: | gap of unknown length        |
| * | 118950 | 119638: | contig of 689 bp in length   |
| * | 119639 | 119718: | gap of unknown length        |
| * | 119719 | 120490: | contig of 772 bp in length   |
| * | 120491 | 120570: | gap of unknown length        |
| * | 120571 | 121288: | contig of 718 bp in length   |
| * | 121289 | 121368: | gap of unknown length        |
| * | 121369 | 122040: | contig of 672 bp in length   |
| * | 122041 | 122120: | gap of unknown length        |
| * | 122121 | 122701: | contig of 581 bp in length   |
| * | 122702 | 122781: | gap of unknown length        |
| * | 122782 | 123465: | contig of 684 bp in length   |
| * | 123466 | 123545: | gap of unknown length        |
| * | 123546 | 124188: | contig of 643 bp in length   |

|                       |  |                |           |                |
|-----------------------|--|----------------|-----------|----------------|
| Query Match           | 2.7%;  | Score 39.6;    | DB 41;    | Length 130583; |
| Best Local Similarity | 61.8%;   | Pred No. 4.4;  |           |                |
| Matches 63;           | Conservative 0;  | Mismatches 39; | Indels 0; | Gaps 0;        |
| Qy 926                | ttgatcaagaactggcagcttcttcgtggccacgaatgagaatgcccgaataattattagat | 985            |           |                |
| Db 93378              | TTGGCCATGATCGGGGTCCCATTAGCCACAAAAGGTAGGCCGCAAAATGCTCTCAGC      | 93437          |           |                |
| Qy 986                | attggcagtgacgctccgctgagagatgcgacgcaattgca                      | 1027           |           |                |
| Db 93438              | ATCGGGGCATCAAGTTCCTGTCGAGGATGCGCGGAGGGGAGCA                    | 93479          |           |                |
| RESULT 10             |  |                |           |                |
| AE003579/c            | 304383 bp  | DNA            | INV       | 22-MAR-2000    |
| LOCUS                 | AE003579   |                |           |                |
| DEFINITION            | Drosophila melanogaster genomic scaffold 142000013386046       | section 5      |           |                |

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

of 16, complete sequence.  
AE003579 AE002638  
AE003579.1 GI:7295765

HTG.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 304383)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,

Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,

George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,

Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,

Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,

Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor

Miklos,G.L., Abriil,J.P., Agbayani,A., An,H.J.,

Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,

Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,

Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,

Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,

Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,

Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de

Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,

Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,

Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,

Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S.,

Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,

Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,

Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,

Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,

Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,

Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,

Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,

McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Murphy,C.,

Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,

Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,

Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,

Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,

Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,

Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,

Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,

Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,

Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,

Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,

Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,

Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,

Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of Drosophila melanogaster

Science 287 (5461), 2185-2195 (2000)

2196006

2 (bases 1 to 304383)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

Location/Qualifiers

1. 304383

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/chromosome="2L"

complement(<9367..>9540)

/gene="CG15419"

/product="CT35478"

/db\_xref="FLYBASE:FBan0015419"

/db\_xref="FLYBASE:FBgn0040710"

/evidence=not\_experimental

complement(<9367..>9540)

/gene="CG15419"

/db\_xref="FLYBASE:FBan0015419"

/db\_xref="FLYBASE:FBgn0040710"

/evidence=not\_experimental

complement(9367..9540)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

mrna

gene

CDS

/ gene="CG15419"
/ note="CG15419 gene product"
/ codon\_start=1
/ db\_xref="FLYBASE:FBan0015419"
/ db\_xref="FLYBASE:FBgn0040710"
/ evidence=not\_experimental
/ protein\_id="AAF51068.1"
/ db\_xref="GI:7295766"
/ translation="MTASGIGTTSATGATSAPOATPLPSIAVSTTASVKGDLGISTT
TTAETSKAFL"
complement(join(<10708..10730,12622..12750,13130..13733,
13815..14033,14369..14563,14624..14741,14797..>15005))
/ gene="Shaw"
/ product="CT9622"
/ db\_xref="FLYBASE:FBan0002822"
/ db\_xref="FLYBASE:FBgn0003386"
/ db\_xref="FLYBASE:FBgn0003386"
complement(<10708..>15005)
/ gene="Shaw"
/ note="CG2822"
/ map="24B-24C"
/ db\_xref="FLYBASE:FBan0002822"
/ db\_xref="FLYBASE:FBgn0003386"
complement(join(10708..10730,12622..12750,13130..13733,
13815..14033,14369..14563,14624..14741,14797..15005))
/ gene="Shaw"
/ note="Shaw gene product"
/ codon\_start=1
/ db\_xref="FLYBASE:FBan0002822"
/ db\_xref="FLYBASE:FBgn0003386"
/ protein\_id="AAF51069.1"
/ db\_xref="GI:7295767"
/ translation="MNLIMDSNRVVLNCGIRHETKATKKIPATRLSRLEALA
NYDPILNEYDFDRPGVFAOVLYNYRTGKLHPDVCGLPEFEELEFWGLDSNOVEPC
CWMTYTOHRTQETLAVLDLDTKPSSEELARKFGFEEDYKGTISWQEMKPR
WSLDEPSSNAKTCIGVSVEFCISILSECLKTHPDMVPIVNTITKTANGSNGW
FLDKTQNAHIAFYICVGNAMFTFEILVRFISPPKNWEFKISSNIDYIATLSY
IDLVLQASHLENADILEFISIRIMRLKLPKTHSSGLKILQTFRASAKELFLLVF
FLVIGVIFASLYVAAEIRIQPNHFNDSIPLGLWALVMTVTGYGDMAPKTYIGMF
VGALCALAGVLTALPVPVIVSNFYNTSHTQARKLPRKRRVLPVEQPRQLPGA
PGVSGCGTGPCSGPHSGPMGSGGTGPRMNNKTKDLYSPKSDMAFSFD"
<18658..>19506
/ gene="CG3410"
/ product="CTL1463"
/ db\_xref="FLYBASE:FBan0003410"
/ db\_xref="FLYBASE:FBgn0031567"
<18658..>19506
/ gene="CG3410"
/ db\_xref="FLYBASE:FBan0003410"
/ db\_xref="FLYBASE:FBgn0031567"
18658..19506
/ gene="CG3410"
/ note="CG3410 gene product"
/ codon\_start=1
/ db\_xref="FLYBASE:FBan0003410"
/ db\_xref="FLYBASE:FBgn0031567"
/ protein\_id="AAF51070.1"
/ db\_xref="GI:7295768"
/ translation="MFRLSVLVNLNLSVHSFAGTAKIEIQPLPALCNGYCFPTLKP
VMEYVAHQDKWNTCTEILANEARKDQIQNLQDLKADVSNIKASQSKDEKLDRM
EREOFAMHSELTNRYITVKLDRTKLQLEAIKNTMDYMKAMDGYFSAINQVCLQIP
GFEKIDGRFYFIEDVELNWLDAQKRRMGCHLASIKTQEFDAIVKELDKSKSYEL
GVNENTKTGFVSAASGKSLYHEWGPGEPHHNDQERCVSILRKLHVNGCTYKRF
ICQYGI"
complement(join(19878..19997,20083..20229,20282..21739,
21895..22294,22650..>22974))
/ gene="CG2818"
/ product="CT9610"
/ db\_xref="FLYBASE:FBan0002818"
/ db\_xref="FLYBASE:FBgn0031566"
complement(<19878..>22974)
/ gene="CG2818"
/ db\_xref="FLYBASE:FBan0002818"
/ db\_xref="FLYBASE:FBgn0031566"

CDS
complement(join(19878..19997,20083..20229,20282..21739,
21895..22293))
/ gene="CG2818"
/ note="CG2818 gene product"
/ codon\_start=1
/ db\_xref="FLYBASE:FBan0002818"
/ db\_xref="FLYBASE:FBgn0031566"
/ protein\_id="AAF51071.1"
/ db\_xref="GI:7295769"
/ translation="MHNWLSHNGDEDAAPSSCPASSENGSRLESRRIFRGDITNRR
LWTFVRLMNDLAPNEQLAIVGCEALGNWQHSAAILSKNEEDNYSNLWTGEIIVP
RHCDETRTAVCAVDEFTQELIVRWETQLQPRIVKELDQPAKNQSDIFGSGNGQEK
WDGWTJTKETLVQLKEFYAFTFKRMKRHIQVQVTPMNLISIPSGAGCDTLPSSPME
DLSNDTHDTKENGSGSTAFSEVYVTSADCEIRSQDQFGTGGPTDLVIFHLTV
GDPENTAYLIDLYSYSRVAKEDGPNHLYHYVLPNLFKRSEGNLELPTTCAKHRP
LGMRLGYLIVKPSLALMDMSVARYWNKWTGLDVGHRSQGTSTFKAKDAVIREN
TITSLANAEHGAQDMVEFDVQLSKDLVPVYHDFMYVLSLKSCLQEHDFLALPMRE
LSLEQLAKLYYHIAEGLSRETRSFHNDCLQHQPFPQLCDVLDALDVHVGFIETKW
SQLEDGKMEEEFEHVVDNRNLYIDCILDVILRRAGNRIVLSFDDPDCITILRFKQNR
YPYMFLLTGRTRTKYOKYMPRGSMELVAVHAMEFLVGVATHEDDLRLDPQOVNLAK
ERGLVLCWGDNDNSKDTIKLKLGLHAIYIDMDVLTTKYKQSFVHQAQDSQKE
LLKLOALEMGHVHNTSADGNEQQA"
join(23524..23991,24270..24433,24726..25649,25705..26284,
26341..26533,26636..26752,26809..26911,26971..27072,
27211..27408,27720..27895,27955..28088,28151..28371,
28807..28870,29245..29464,30470..30624,30795..30916,
31389..31506,31753..>31967)
/ gene="CG10030"
/ product="CT28231"
/ db\_xref="FLYBASE:FBan0010030"
/ db\_xref="FLYBASE:FBgn0031565"
<23524..>31967
/ gene="CG10030"
/ db\_xref="FLYBASE:FBan0010030"
join(23604..23991,24270..24433,24726..25649,25705..26284,
26341..26533,26636..26752,26809..26911,26971..27072,
27211..27408,27720..27895,27955..28088,28151..28371,
28807..28870,29245..29464,30470..30624,30795..30916,
31389..31506,31753..31967)
/ gene="CG10030"
/ note="CG10030 gene product"
/ codon\_start=1
/ db\_xref="FLYBASE:FBan0010030"
/ db\_xref="FLYBASE:FBgn0031565"
/ protein\_id="AAF51072.1"
/ db\_xref="GI:7295770"
/ translation="MEGDKVVFHSSDLPYEVLYDDEKHVSHMPDSKLSVRNILDAYA
SAGKILITRKLRSIPYPHARVKNIGIRVYKSEPRKQSYKSSLWKEAYNVRNRR
HAGHFTCFISDTPPAVCYCRANRSKGTPIKREPMCSRLIINDDFEIRNAKNTH
FRHFHFGSLSVNKEEETRKFILKLVGSRFSTSMQYDPDENEFEFLQYQDELE
LLEDLPDEFNAYDGPSTSKQAAEKQENRAPVAALRSTRLGNSTLGSQPSQLSTFGA
Query Match 2.7%; Score 39.6; DB 34; Length 304383;
Best Local Similarity 61.8%; Pred. No. 5.2;
Matches 63; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
Qy 926 ttgatcaagaactggcagcttctccctggccacgaatggagaatcgcgaataattat985
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140969 TTGGCCATGATCTGGGGGTCCCATTAGCCACAAAAGGTAGCGCGGAAATCTCTCAGC 140910
Qy 986 attggcagtcgacgtcccgctgaggatgcgacgcgaattgca 1027
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140909 ATCGGGGCATCAAGTTCGTCGATCGATCGCGGAGGGAGCA 140868
RESULT 11
HS288L1/c 118593 bp DNA PRI 12-DEC-1999
LOCUS Human DNA sequence from clone RP1-288L1 on chromosome 22 Contains
DEFINITION an STSs, GSSs and genomic marker D22S1152, complete sequence.
ACCESSION 282196
VERSION 282196.2 GI:6572205

KEYWORDS HTG: D22S1152.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 118593)  
AUTHORS Burton,J.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerquest@sanger.ac.uk  
COMMENT On Dec 13, 1999 this sequence version replaced gi:1903194.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
RP1-288L1 is from the library RPCI-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP1-288L1.  
FEATURES  
Location/Qualifiers  
1..118593  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="RP1-288L1"  
/clone\_lib="RPCI-1"  
4..141  
/note="WSTA repeat: matches 185..314 of consensus"  
repeat\_region 142..852  
/note="L1PA8 repeat: matches 534..6163 of consensus"  
repeat\_region 853..1031  
/note="WSTA repeat: matches 1..200 of consensus"  
repeat\_region 1037..1809  
/note="HERVL repeat: matches 3871..4641 of consensus"  
repeat\_region 2248..2413  
/note="HERVL repeat: matches 3359..3525 of consensus"  
repeat\_region 3173..3373  
/note="LTR33 repeat: matches 8..223 of consensus"  
repeat\_region 3673..4275  
/note="L1M4 repeat: matches 2681..3306 of consensus"  
repeat\_region 4325..4939  
/note="HERVL16 repeat: matches 1665..2289 of consensus"  
repeat\_region 4975..5684  
/note="HERVL16 repeat: matches 2344..3072 of consensus"  
repeat\_region 5924..6339  
/note="HERVL repeat: matches 3442..3860 of consensus"  
misc\_feature 6296..6578  
/note="match: STS: Em:F08316"  
repeat\_region 6452..6536  
/note="HERVL repeat: matches 4031..4115 of consensus"  
repeat\_region 6537..6824  
/note="AluSc repeat: matches 13..299 of consensus"  
repeat\_region 6825..7958  
/note="HERVL repeat: matches 4115..5247 of consensus"

repeat\_region 8198..8219  
/note="11 copies 2 mer aa 100 conserved"  
repeat\_region 8803..8918  
/note="L2 repeat: matches 2624..2748 of consensus"  
repeat\_region 9068..9178  
/note="MER94 repeat: matches 1..113 of consensus"  
repeat\_region 9957..15197  
/note="L1PA2 repeat: matches 900..6146 of consensus"  
repeat\_region 15193..15979  
/note="L1PA2 repeat: matches 1..776 of consensus"  
repeat\_region 16282..17315  
/note="SVA repeat: matches 14..955 of consensus"  
misc\_feature 16843..17747  
/note="Cpg island"  
/evidence="not\_experimental"  
repeat\_region 17316..17875  
/note="SVA repeat: matches 798..1363 of consensus"  
misc\_feature complement(17752..18176)  
/note="match: GSS: Em:AQ113498"  
misc\_feature 17896..18316  
/note="match: GSS: Em:AQ802108"  
repeat\_region 18158..18396  
/note="MIR repeat: matches 15..262 of consensus"  
repeat\_region 18529..18648  
/note="L2 repeat: matches 2630..2749 of consensus"  
repeat\_region 18682..18805  
/note="MIR repeat: matches 52..173 of consensus"  
repeat\_region 19153..19462  
/note="AluX repeat: matches 1..310 of consensus"  
repeat\_region 20064..20128  
/note="L1PA4 repeat: matches 6075..6139 of consensus"  
repeat\_region 20456..20899  
/note="L2 repeat: matches 2218..2742 of consensus"  
repeat\_region 21834..23162  
/note="L1MC2 repeat: matches 4679..6064 of consensus"  
repeat\_region 23163..23463  
/note="AluSg repeat: matches 1..302 of consensus"  
repeat\_region 23464..23735  
/note="L1MC2 repeat: matches 6064..6328 of consensus"  
repeat\_region 23738..23779  
/note="21 copies 2 mer aa 76 conserved"  
repeat\_region 24053..24998  
/note="L1MA8 repeat: matches 5344..6264 of consensus"  
repeat\_region 25021..25378  
/note="L2 repeat: matches 1660..2020 of consensus"  
repeat\_region 25440..25665  
/note="MIR repeat: matches 8..243 of consensus"  
repeat\_region 26462..26542  
/note="MIR repeat: matches 52..147 of consensus"  
repeat\_region 26815..27048  
/note="MIR repeat: matches 13..251 of consensus"  
repeat\_region 27272..27533  
/note="AluSg repeat: matches 1..264 of consensus"  
repeat\_region 28220..28347  
/note="MIR repeat: matches 18..152 of consensus"  
repeat\_region 28620..28790  
/note="MIR repeat: matches 89..261 of consensus"  
repeat\_region 28835..28983  
/note="MIR repeat: matches 90..231 of consensus"  
repeat\_region 29619..29735  
/note="L2 repeat: matches 2418..2529 of consensus"  
repeat\_region 31018..31154  
/note="L2 repeat: matches 2240..2400 of consensus"  
repeat\_region 31217..31587  
/note="THE1C repeat: matches 1..371 of consensus"  
repeat\_region 31723..31850  
/note="L2 repeat: matches 2577..2707 of consensus"  
repeat\_region 32234..32433  
/note="MIR repeat: matches 47..256 of consensus"  
repeat\_region 32507..32734  
/note="MIR repeat: matches 28..262 of consensus"  
repeat\_region 33524..33626  
/note="MIR repeat: matches 23..121 of consensus"

**Tue Sep 12 08:47:29 2000**

```

repeat_region 33957. .34137
/note="MIR repeat: matches 52. .262 of consensus"
repeat_region 34184. .34257
/note="I2 repeat: matches 2576. .2646 of consensus"
repeat_region 34305. .34649
/note="Alusq repeat: matches 1. .309 of consensus"
repeat_region 35067. .35231
/note="MIR repeat: matches 67. .240 of consensus"
repeat_region 35234. .35296
/note="I2 repeat: matches 2647. .2709 of consensus"
repeat_region 35955. .36499
/note="MER69 repeat: matches 2000. .2510 of consensus"
repeat_region 36499. .37292
/note="MER69 repeat: matches 1. .801 of consensus"
repeat_region 37882. .37967
/note="Alusx repeat: matches 9. .294 of consensus"
repeat_region 38671. .38783
/note="MIR repeat: matches 22. .145 of consensus"
repeat_region 39632. .40082
/note="MER70A repeat: matches 136. .593 of consensus"
repeat_region 40083. .40294
/note="MIR repeat: matches 49. .261 of consensus"
repeat_region 40320. .40453
/note="I2 repeat: matches 2616. .2749 of consensus"
repeat_region 40621. .40689
/note="I2 repeat: matches 2592. .2673 of consensus"
repeat_region 41643. .41745
/note="MIR repeat: matches 48. .145 of consensus"
repeat_region 42285. .42925
/note="MER1B repeat: matches 1. .634 of consensus"
repeat_region 42921. .43422
/note="MLTIG repeat: matches 2. .512 of consensus"
repeat_region 43464. .44153
/note="HERVL repeat: matches 4683. .5399 of consensus"
/note="match: GSS: Em:A0126732"
misc_feature complement(44300. .44703)
/note="match: GSS: Em:A0790304"
misc_feature 44690. .44996
/note="match: GSS: Em:A0618052"
repeat_region 44839. .45277
/note="I2 repeat: matches 2293. .2749 of consensus"
repeat_region 45279. .45426
/note="MIR repeat: matches 87. .243 of consensus"
repeat_region 47053. .47490
/note="MLTIC repeat: matches 1. .465 of consensus"
misc_feature 48174. .48491
/note="Tandem repeat"
repeat_region 48187. .48488

Query Match 2.7%; Score 39.4; DB 10; Length 118593;
Best Local Similarity 51.4%; Pred. No. 5;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps

QY 160 tgaactaaacgcgtataaaattacgtggagaataacatatgaatggttggagaagaatccat 219
Db 94668 TTTACCTAAACTCTCTAGTGTCTATGTTGTTTAACTCTAAAATGAAGTCAGAAATACAC 94609

QY 220 tatggccaaacgcggtgtgttggtccggggcgtaaacgcgttaacgcgtacacctgacggaaga 279
Db 94608 TATATCCACCTCATAGCAGCGTGGGCGCTGCCAGTAAATACATCACTGGATATGCA 94549

QY 280 aatgcaaaagagtttcattacaccattggcccttattccacaccgctcctgacct 336

```

|            |  |
|------------|--|
| RESULT     | 12   |
| AE002546   |  |
| LOCUS      | AE002546 10619 bp DNA BCT  |
| DEFINITION | Neisseria meningitidis serogroup B strain MC58 section 188 of 206<br>of the complete genome. |
| ACCESSION  | AE002546 AE002098  |

AE002546.1 GI:7227234  
 .  
 Neisseria meningitidis.  
 Neisseria meningitidis  
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
 Neisseria.  
 1 (bases 1 to 10619)  
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,  
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,  
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,  
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,  
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,  
 Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,  
 Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,  
 Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,  
 Rappuoli,R. and Venter,J.C.  
 Complete genome sequence of *Neisseria meningitidis* serogroup B  
 strain MC58  
 Science 287 (5459), 1809-1815 (2000)  
 20175755  
 2 (bases 1 to 10619)  
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,  
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,  
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,  
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,  
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,  
 Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,  
 Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,  
 Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,  
 Rappuoli,R. and Venter,J.C.  
 Direct Submission  
 Submitted (17-Mar-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source  
 .  
 /organism="Neisseria meningitidis"  
 /strain="MC58"  
 /db\_xref="taxon:487"  
 /note="serogroup: B"  
 complement(77..367)  
 /gene="NMBl973"  
 complement(77..367)  
 /gene="NMBl973"  
 /note="similar to SP:P77913 PID:1513097 percent identity:  
 98.96; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chaperonin, 10 kDa"  
 /protein\_id="AA042302.1"  
 /db\_xref="GI:7227235"  
 /translation="MTIRPLHDRVVVKRLAEAEKTAGSIVLPGAAAEKPDGMGEIVAG  
 AGKIGKDGSRRLDQIKVGDKIIIFGYSQGTAKDGBELLVMEEDIFGIVEK"  
 595..1247  
 /gene="NMBl974"  
 /note="This region contains a gene with one or more  
 premature stops or frameshifts, and is not the result of a  
 sequencing artifact; similar to GB:X59756 percent  
 identity: 82.61; identified by sequence similarity;  
 putative:IS1016C2 transposase, degenerate"  
 1460..2995  
 /gene="NMBl975"  
 1460..2995  
 /gene="NMBl975"  
 /note="similar to GB:L42023 SP:P44849 PID:1005680  
 PID:1220824 PID:1204983 percent identity: 77.47;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="sodium- and chloride-dependent transporter"  
 /protein\_id="AA042303.1"  
 /db\_xref="GI:7227236"  
 /translation="MSDSKTKERATFGTRAFMAAIGSAVLGNWRFPYIAFENG  
 GAFILPYALLATGPIELLIDIAIGHRYRGSAPLAFRLGRWFEPVGNWNTNIV



AC015424  
 VERSION AC015424.1 GI:6435911  
 HTG: HTGS\_PHASE2  
 SOURCE fruit fly  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila  
 1 (bases 1 to 133144)  
 AUTHORS Adams, M. and Venter, J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT This sequence was identified as CDM:10214369 by the submitter.  
 For further information on this sequence e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* this sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES  
 source  
 1..133144  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 BASE COUNT 41011 a 26133 c 25910 g 40090 t  
 ORIGIN  
 Query Match 2.78; Score 39.2; DB 42; Length 133144;  
 Best Local Similarity 49.18; Pred. No. 5.8;  
 Matches 104; Conservativity 0; Mismatches 108; Indels 0; Gaps 0;  
 QY 320 acaccgtcctgaccatcgaccgcggtagccggtattgtcgacactcgagatgctttt 379  
 Db 17762 ACCACCGGGTGGTATCGGTATGGCAGACTGATTAATGCTGCTCGAGTGGCTAAT 17703  
 QY 380 gaaggtgctatcaatcggaacaggatattccagccagtgtgtaaaatgcccttttc 439  
 Db 17702 GGACTCTTCTTCTCGGGGGGATGGCTCATTTGAACCTCTCGAGACTCCCGGCC 17643  
 QY 440 aaccacaaacgacgcgatcattggtcgaatgcccggagaaaggatgctgcctctc 499  
 Db 17642 GCGCGTAAGTTTCTGATAGCGATCGCAACCTGCTCCCAAGTGGCTCATGATCGCTCTG 17583  
 QY 500 tatatgaatccatgttgcgcgcgcgcgttga 531  
 Db 17582 GTACTCGCGCTCATCTCGCCCGCTCGAGGTCA 17551  
 RESULT 14  
 AC010003/c  
 LOCUS  
 DEFINITION Drosophila melanogaster clone RPC198-6L11, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 70 unordered pieces.  
 AC010003  
 VERSION AC010003.5 GI:6996673  
 HTG: HTGS\_PHASE1  
 SOURCE fruit fly  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila  
 1 (bases 1 to 133190)  
 AUTHORS Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,  
 Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
 Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
 David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
 Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,  
 Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,  
 Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M.,  
 Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, F., Jones, M.,  
 Kelly, S., Kondajewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,  
 Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,

Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,  
 Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,  
 Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R.,  
 Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahbah, M.,  
 Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,  
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and  
 Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 133190)  
 AUTHOR Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-SEP-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Feb 18, 2000 this sequence version replaced gi:5902251.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: DRAK  
 Center clone name: RPC198-6L11  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-terminator Big Dye; 15% of reads  
 Assembly program: Phrap; version 0.980611  
 Consensus quality: 77406 bases at least Q40  
 Consensus quality: 98895 bases at least Q30  
 Consensus quality: 108419 bases at least Q20  
 Estimated insert size: 120294; sum-of-contigs estimation  
 Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 70 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 783 782: contig of 782 bp in length  
 803 802: gap of unknown length  
 1638 1637: contig of 835 bp in length  
 1658 1657: gap of unknown length  
 2973 2972: contig of 1315 bp in length  
 4108 4107: gap of unknown length  
 4128 4127: contig of 1115 bp in length  
 5425 5424: gap of unknown length  
 6127 6126: gap of unknown length  
 7389 7388: contig of 682 bp in length  
 7409 7408: contig of 1242 bp in length  
 8725 8724: gap of unknown length  
 8745 8744: gap of unknown length  
 9551 9550: contig of 806 bp in length  
 9571 9570: gap of unknown length  
 10385 10384: contig of 815 bp in length  
 10406 10405: gap of unknown length  
 11616 11615: contig of 1210 bp in length  
 11636 11635: gap of unknown length  
 12454 12453: contig of 818 bp in length  
 12474 12473: gap of unknown length  
 13165 13164: contig of 691 bp in length  
 13185 13184: gap of unknown length  
 14008 14007: contig of 823 bp in length  
 14028 14027: gap of unknown length  
 15565 15565: contig of 1538 bp in length



15566 15585: gap of unknown length  
15586 16369: contig of 984 bp in length  
15589 16589: gap of unknown length  
16590 17562: contig of 973 bp in length  
17563 17582: gap of unknown length  
17583 18414: contig of 832 bp in length  
18415 18434: gap of unknown length  
18435 19614: contig of 1180 bp in length  
19615 19634: gap of unknown length  
20885 20885: contig of 1251 bp in length  
20905 20905: gap of unknown length  
20906 22136: contig of 1231 bp in length  
22137 22156: gap of unknown length  
22157 23503: contig of 1347 bp in length  
23504 23524: gap of unknown length  
23524 24606: contig of 1083 bp in length  
24607 24626: gap of unknown length  
25765 25765: contig of 1139 bp in length  
25766 25785: gap of unknown length  
25786 26995: contig of 1110 bp in length  
26996 26915: gap of unknown length  
26916 28149: contig of 1234 bp in length  
28150 28169: gap of unknown length  
28170 29503: contig of 1334 bp in length  
29504 29523: gap of unknown length  
29524 31129: contig of 1606 bp in length  
31130 31149: gap of unknown length  
31150 32506: contig of 1357 bp in length  
32507 32526: gap of unknown length  
32527 33151: contig of 625 bp in length  
33152 33171: gap of unknown length  
33172 34833: contig of 1662 bp in length  
34834 34853: gap of unknown length  
34854 36463: contig of 1610 bp in length  
36464 36483: gap of unknown length  
36484 37769: contig of 1286 bp in length  
37770 37789: gap of unknown length  
37790 39247: contig of 1458 bp in length  
39248 39267: gap of unknown length  
39268 41092: contig of 1825 bp in length  
41093 41112: gap of unknown length  
41113 42002: contig of 890 bp in length  
42003 42022: gap of unknown length  
42023 43449: contig of 1627 bp in length  
43450 43669: gap of unknown length  
43670 45468: contig of 1799 bp in length  
45469 45488: gap of unknown length  
45489 46976: contig of 1488 bp in length  
46977 46996: gap of unknown length  
46997 48341: contig of 1345 bp in length  
48342 48361: gap of unknown length  
48362 50274: contig of 1913 bp in length  
50275 50294: gap of unknown length  
50295 52835: contig of 2541 bp in length  
52836 52855: gap of unknown length  
52856 54462: contig of 1607 bp in length  
54463 54482: gap of unknown length  
54483 55386: contig of 904 bp in length  
55387 55406: gap of unknown length  
55407 57489: contig of 2083 bp in length  
57490 57509: gap of unknown length  
60192 60192: contig of 2683 bp in length  
60193 60212: gap of unknown length  
60213 62373: contig of 2161 bp in length  
62374 62393: gap of unknown length  
62394 64223: contig of 1830 bp in length  
64224 64243: gap of unknown length  
64244 66061: contig of 1818 bp in length  
66062 68081: gap of unknown length  
68082 68233: contig of 2152 bp in length  
68234 68253: gap of unknown length  
68254 70076: contig of 1823 bp in length  
70077 70096: gap of unknown length

70097 72046: contig of 1950 bp in length  
72047 72066: gap of unknown length  
72067 74963: contig of 2897 bp in length  
74964 74983: gap of unknown length  
74984 77410: contig of 2427 bp in length  
77411 77430: gap of unknown length  
77431 80001: contig of 2571 bp in length  
80002 80021: gap of unknown length  
80022 81918: contig of 1897 bp in length  
81919 81938: gap of unknown length  
81939 83880: contig of 1942 bp in length  
83881 83900: gap of unknown length  
83901 87305: contig of 3405 bp in length  
87306 87325: gap of unknown length  
87326 90628: contig of 3303 bp in length  
90629 90648: gap of unknown length  
90649 93277: contig of 2629 bp in length  
93278 93297: gap of unknown length  
93298 95346: contig of 2029 bp in length  
95347 95366: gap of unknown length  
95367 98750: contig of 3404 bp in length  
98751 98770: gap of unknown length  
98771 101763: contig of 2993 bp in length  
101764 101783: gap of unknown length  
101784 105858: contig of 4075 bp in length  
105859 105878: gap of unknown length  
105879 108448: contig of 2570 bp in length  
108449 108468: gap of unknown length  
108469 112321: contig of 3853 bp in length  
112322 112341: gap of unknown length  
112342 116841: contig of 4500 bp in length

Query Match 2.7%; Score 39.2; DB 51; Length 133190;  
Best Local Similarity 49.1%; Pred. No. 5.8;  
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 320 acaccgcctgaccatcgaaacccggtgaccggtatttcgacacactgagatgctttt 379  
Db 111107 ACCACCGGGTGGCTATCGGTATGGCAGACTGGATTATTTCGTGCTGCGATGCTATT 111048

QY 380 gaagggtctaatcgcgaacaggtattccgagccagtgctctaaagcccttctc 439  
Db 111047 GGAGCTCTTCTTCTTCGGGGGGGATGCGTCCATTGAACCTCCGAGACGCTCCCGGCC 110988

QY 440 aaccacaacacgacgatcgatggtcaatggcgcggagagaagtgatgctgcgtgc 499  
Db 110987 GGCGGTAAAGTTCTGCTAGGATCGCCAAACTCGTCCCAAGTGGCCCTCATGATGCTCTG 110928

QY 500 tatatcgatccatgttgcgccgcgcgttga 531  
Db 110927 GTACTCGGCTCATCTCTCGCCGCTCGAGGTCA 110896

RESULT 15  
AC009369/c

LOCUS 154381 bp DNA 18-FEB-2000  
DEFINITION Drosophila melanogaster clone RPC198-8G12, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 74 unordered pieces.  
ACCESSION AC009369  
VERSION AC009369.5 GI:6996666  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 154381)  
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Doman-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,

Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,  
 Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
 Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,  
 Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,  
 Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,  
 Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,  
 Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,  
 Tabors, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahbah, M.,  
 Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,  
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and  
 Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 154381)  
 Worley, K.C.  
 Submitted (20-AUG-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Bay Plaza, Houston, TX 77030, USA  
 On Feb 18, 2000 this sequence version replaced gi:5881432.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: DRAD  
 Center clone name: RPC198-8G12  
 ----- Summary Statistics  
 Sequencing vector: M13, L08821  
 Chemistry: Dye-terminator Big Dye: 4% of reads  
 Assembly program: Phrap: version 0.980611  
 Consensus quality: 100833 bases at least Q40  
 Consensus quality: 122674 bases at least Q30  
 Consensus quality: 132575 bases at least Q20  
 Estimated insert size: 144035; sum-of-contigs estimation  
 Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 74 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 782: contig of 782 bp in length  
 \* 803: gap of unknown length  
 \* 1639: contig of 836 bp in length  
 \* 1659: gap of unknown length  
 \* 2547: contig of 888 bp in length  
 \* 2567: gap of unknown length  
 \* 3736: contig of 1169 bp in length  
 \* 3756: gap of unknown length  
 \* 4642: contig of 887 bp in length  
 \* 4643: gap of unknown length  
 \* 4663: contig of 921 bp in length  
 \* 5584: gap of unknown length  
 \* 5604: contig of 1250 bp in length  
 \* 6854: gap of unknown length  
 \* 6874: contig of 1022 bp in length  
 \* 7896: gap of unknown length  
 \* 7916: contig of 1022 bp in length  
 \* 8937: gap of unknown length  
 \* 8958: contig of 1582 bp in length  
 \* 10540: gap of unknown length  
 \* 10560: contig of 1343 bp in length  
 \* 11903: gap of unknown length  
 \* 11923: contig of 1522 bp in length  
 \* 13445: gap of unknown length

13465  
 14837  
 14856: gap of unknown length  
 16113: contig of 1257 bp in length  
 16114: gap of unknown length  
 16134: contig of 820 bp in length  
 16953: gap of unknown length  
 16973: gap of unknown length  
 17949: contig of 976 bp in length  
 17950: gap of unknown length  
 17951: contig of 1282 bp in length  
 19251: gap of unknown length  
 19271: contig of 876 bp in length  
 20147: gap of unknown length  
 20167: gap of unknown length  
 21490: contig of 1323 bp in length  
 21510: gap of unknown length  
 22371: contig of 861 bp in length  
 22391: gap of unknown length  
 23573: contig of 1182 bp in length  
 23593: gap of unknown length  
 24778: contig of 1185 bp in length  
 24798: gap of unknown length  
 26680: contig of 1882 bp in length  
 26700: gap of unknown length  
 27675: contig of 975 bp in length  
 27695: gap of unknown length  
 28733: contig of 1038 bp in length  
 28753: gap of unknown length  
 30178: contig of 1425 bp in length  
 30198: gap of unknown length  
 31626: contig of 1428 bp in length  
 31646: gap of unknown length  
 32836: contig of 1190 bp in length  
 32856: gap of unknown length  
 34558: contig of 1702 bp in length  
 34578: gap of unknown length  
 35924: contig of 1346 bp in length  
 35944: gap of unknown length  
 37736: contig of 1792 bp in length  
 37756: gap of unknown length  
 38769: contig of 1013 bp in length  
 38789: gap of unknown length  
 39846: contig of 1057 bp in length  
 39866: gap of unknown length  
 41513: contig of 1647 bp in length  
 41533: gap of unknown length  
 43110: contig of 1577 bp in length  
 43130: gap of unknown length  
 45511: contig of 2381 bp in length  
 45531: gap of unknown length  
 46655: contig of 1124 bp in length  
 46675: gap of unknown length  
 47948: contig of 1273 bp in length  
 47968: gap of unknown length  
 50813: contig of 2845 bp in length  
 50833: gap of unknown length  
 53480: contig of 2647 bp in length  
 53500: gap of unknown length  
 54934: contig of 1434 bp in length  
 54954: gap of unknown length  
 57412: contig of 2458 bp in length  
 57432: gap of unknown length  
 59817: contig of 2385 bp in length  
 59837: gap of unknown length  
 61341: contig of 1504 bp in length  
 61361: gap of unknown length  
 63044: contig of 1683 bp in length  
 63064: gap of unknown length  
 64833: contig of 1769 bp in length  
 64853: gap of unknown length  
 67374: contig of 2521 bp in length  
 67394: gap of unknown length  
 68813: contig of 1419 bp in length  
 68833: gap of unknown length  
 70648: contig of 1815 bp in length

```
* 70649 70668: gap of unknown length
* 70669 71950: contig of 1282 bp in length
* 71951 71970: gap of unknown length
* 71971 74720: contig of 2750 bp in length
* 74721 74740: gap of unknown length
* 74741 76716: contig of 1976 bp in length
* 76717 76736: gap of unknown length
* 76737 78617: contig of 1881 bp in length
* 78618 78637: gap of unknown length
* 78638 81017: contig of 2380 bp in length
* 81018 81037: gap of unknown length
* 81038 83770: contig of 2733 bp in length
* 83771 83790: gap of unknown length
* 83791 86125: contig of 2335 bp in length
* 86126 86145: gap of unknown length
* 86146 89229: contig of 3084 bp in length
* 89230 89249: gap of unknown length
* 89250 90667: contig of 1418 bp in length
* 90668 90687: gap of unknown length
* 90688 92693: contig of 2006 bp in length
* 92694 92713: gap of unknown length
* 92714 95077: contig of 2364 bp in length
* 95078 95097: gap of unknown length
* 95098 97907: contig of 2810 bp in length
* 97908 97927: gap of unknown length
* 97928 100693: contig of 2766 bp in length
* 100694 100713: gap of unknown length
* 100714 103619: contig of 2906 bp in length
* 103620 103639: gap of unknown length
* 103640 106885: contig of 3246 bp in length
* 106886 106905: gap of unknown length
* 106906 109193: contig of 2288 bp in length
* 109194 109213: gap of unknown length
* 109214 112158: contig of 2945 bp in length
* 112159 112178: gap of unknown length
* 112179 116985: contig of 4807 bp in length

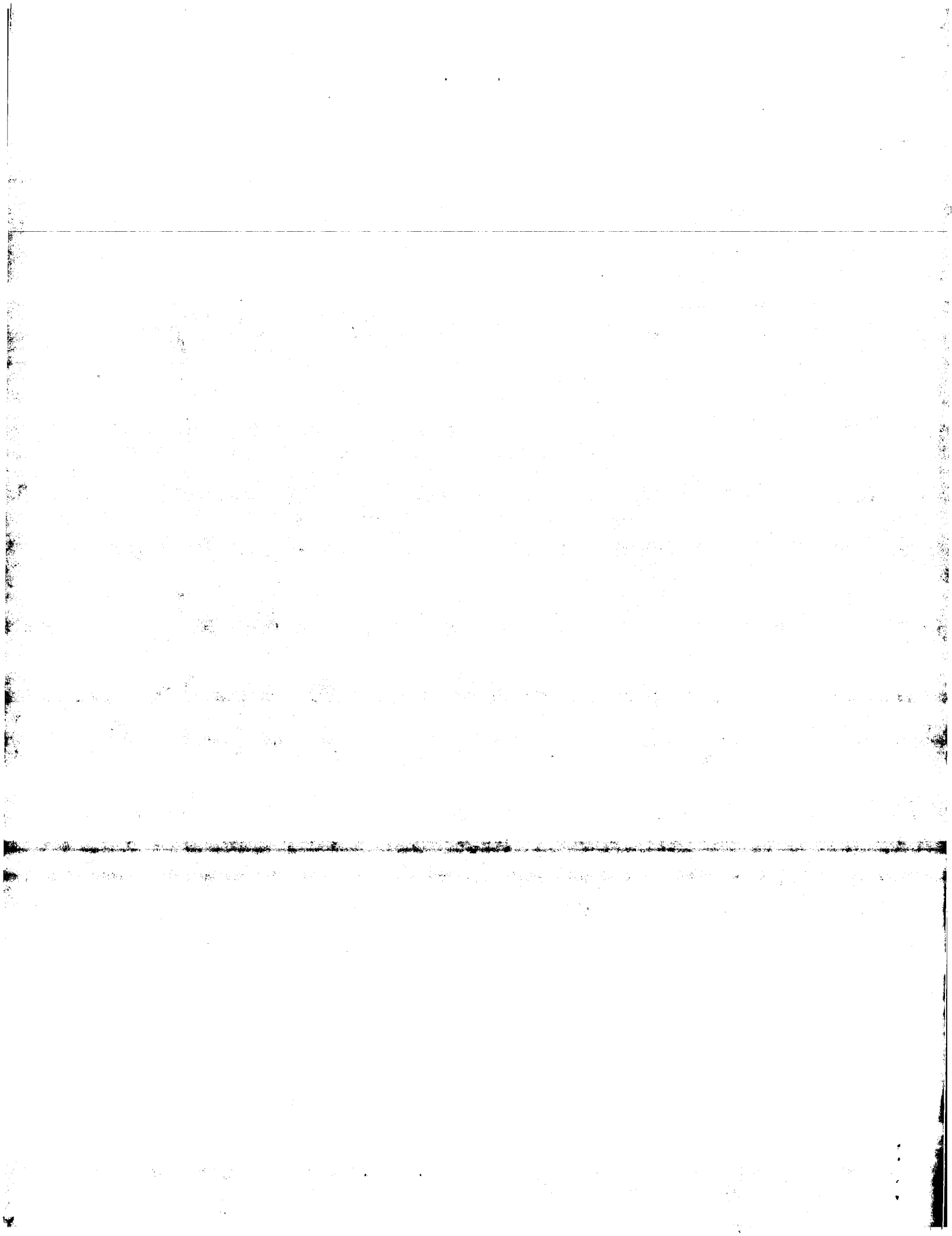
Query Match      2.7%; Score 39.2; DB 51; Length 154381;
Best Local Similarity 49.1%; Pred. No. 6;
-Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 320 acaccgctcctgaccatcgaccggtgaccgattattgtcgacactcgagatgctttt 379
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140033 ACCACCGGGCTGCTATCGGTATGCGAGACTGGATTATTGCTGCTCGAGTGGCTATT 139974
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 380 gaagggtgatcaattcggaaacaggattatccgagcagtgctaaaaatgccctttctc 439
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139973 GGAGCTCTTCTTCTTCGGGGCGGGATGCGTCCATTGCACTCTTCGACAGCTCCCGGCC 139914
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 440 aaccacaaacgaccgatcatggtcaatggcgcgagagaaagtgatgctcgctgtc 499
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139913 GGCCGTAAGTTGTGCTAGCGATCGCCAAACTCGTCCCAAGTGGCCCTCATGCTGCTCG 139854
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 500 tatatgaatccatgttgcccccgcggttga 531
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139853 GTACTCGCTCATCTCTCGCCCGTCGAGGTCA 139822
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:18:12 ; Search time 67.18 Seconds  
(without alignments)  
5370.303 Million cell updates/sec

Title: US-09-214-679-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtggcc.....aatgcaattcatttgatcc 1442

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description        |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1          | 1442  | 100.0       | 1442   | 1  | V10449   | K. oxytoca R-speci |
| C 2        | 39.2  | 2.7         | 3900   | 1  | O66950   | Cell death hid cDN |
| C 3        | 37.6  | 2.6         | 502    | 1  | Q66948   | Cell death hid gen |
| C 4        | 35.2  | 2.4         | 429    | 1  | X28300   | Human CYP3A4 gene  |
| - 5        | 34.2  | 2.4         | 5733   | 1  | X00477   | Arabidopsis thalia |
| C 6        | 33.2  | 2.3         | 745    | 1  | X14578   | H. pylori GHPO 166 |
| C 7        | 33.2  | 2.3         | 3201   | 1  | O89701   | Alpha-1,4-glucan l |
| C 8        | 33.2  | 2.3         | 3201   | 1  | O87616   | M.costata DNA enco |
| C 9        | 33.2  | 2.3         | 3201   | 1  | R84194   | Morchella costata  |
| C 10       | 32.8  | 2.3         | 383    | 1  | V88779   | EST clone HO266. N |
| C 11       | 32.2  | 2.2         | 1226   | 1  | Q57702   | K.lactis transaldo |
| C 12       | 32.2  | 2.2         | 1349   | 1  | Q57701   | K.lactis transaldo |
| C 13       | 32.2  | 2.2         | 6260   | 1  | V52729   | Human hepatocyte n |
| C 14       | 31.6  | 2.2         | 618    | 1  | Q48718   | Encodes protein ho |
| C 15       | 31.2  | 2.2         | 2167   | 1  | T28490   | P. aeruginosa dete |
| C 16       | 31.2  | 2.2         | 110000 | 1  | V30458_0 | Rhizobium species  |
| C 17       | 31.2  | 2.2         | 110000 | 1  | V30459_0 | Rhizobium species  |
| C 18       | 31.2  | 2.1         | 1443   | 1  | V52023   | Helicobacter polyp |
| C 19       | 30.8  | 2.1         | 1187   | 1  | T39536   | E. coli lacI methy |
| C 20       | 30.8  | 2.1         | 1188   | 1  | T39535   | E. coli lacI methy |
| C 21       | 30.6  | 2.1         | 1542   | 1  | V12274   | Chloramphenicol re |
| C 22       | 30.4  | 2.1         | 1049   | 1  | Q93164   | Neutral protease n |
| C 23       | 30.2  | 2.1         | 10095  | 1  | V58938   | Mycobacterium tube |
| C 24       | 29.8  | 2.1         | 417    | 1  | V53317   | DNA encoding a Sta |
| C 25       | 29.8  | 2.1         | 7494   | 1  | V74380   | Staphylococcus aur |
| C 26       | 29.6  | 2.1         | 11597  | 1  | X13159   | Enterococcus faeca |
| C 27       | 29.4  | 2.0         | 1888   | 1  | V31482   | Papaya ACC synthas |
| C 28       | 29.4  | 2.0         | 2221   | 1  | V65258   | DNA encoding a S.  |
| C 29       | 29.4  | 2.0         | 4665   | 1  | T97955   | Human epididymis-s |
| C 30       | 29.4  | 2.0         | 7902   | 1  | V52245   | Streptococcus pneu |
| C 31       | 29.2  | 2.0         | 1620   | 1  | T67858   | H. pylori cytoplas |
| C 32       | 29.2  | 2.0         | 17250  | 1  | X13020   | Enterococcus faeca |
| C 33       | 29.2  | 2.0         | 46899  | 1  | Q54386   | T. niueum Cyclospo |

|      |      |     |        |   |          |                     |
|------|------|-----|--------|---|----------|---------------------|
| 34   | 29.2 | 2.0 | 110000 | 1 | V30458_4 | Continuation (5 of  |
| 35   | 29.2 | 2.0 | 110000 | 1 | V30459_4 | Continuation (5 of  |
| 36   | 29   | 2.0 | 300    | 1 | T34990   | N. gonorrhoeae-spe  |
| 37   | 29   | 2.0 | 850    | 1 | Q51201   | Sequence specific   |
| 38   | 29   | 2.0 | 850    | 1 | T34986   | Neisseria gonorrhoe |
| 39   | 29   | 2.0 | 1067   | 1 | V33485   | Neisseria gonorrhoe |
| 40   | 29   | 2.0 | 3596   | 1 | O52699   | Corticosterone-bin  |
| 41   | 29   | 2.0 | 25580  | 1 | X12956   | Enterococcus faeca  |
| 42   | 28.8 | 2.0 | 1479   | 1 | X34596   | DNA encoding a ccm  |
| C 43 | 28.6 | 2.0 | 1023   | 1 | X34232   | Mycobacterium spec  |
| C 44 | 28.6 | 2.0 | 2019   | 1 | X19481   | Staphylococcus aur  |
| C 45 | 28.6 | 2.0 | 2350   | 1 | X13046   | Enterococcus faeca  |

ALIGNMENTS

RESULT 1  
ID V10449  
AC V10449  
DT 16-JUN-1998 (first entry)  
DE K. oxytoca R-specific amidohydrolase genomic DNA.  
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;  
OS Klebsiella oxytoca.  
FH Key Location/Qualifiers  
CDS 197..1183  
FT /\*tag= a  
FT /product= /note= "R-specific amidohydrolase"  
FN W09801568-A2.  
PD 15-JAN-1998.  
PF 10-JUL-1997; E03670.  
PR 03-MAR-1997; CH-000500.  
PR 10-JUL-1996; CH-001723.  
PA (LONZ ) LONZA AG.  
PI Brieden W, Naughton A, Robins K, Shaw N, Tinschert A,  
PI Zimmermann T;  
DR WPI: 98-101063/09.  
DR P-PSDB: W40263.  
PT Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation  
PT - by stereoselective hydrolysis of corresponding racemic amide using  
PT microorganism or derived enzyme, used as drug intermediate  
PS Claim 7a; Page 32-34; 68pp; German.  
CC This DNA sequence encodes a R-specific amidohydrolase isolated from  
CC Klebsiella oxytoca strain PRS1 which allows the microorganism to  
CC utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole  
CC nitrogen source. This amidohydrolase is used in a process for preparing  
CC (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which  
CC is cheaper than prior art optical resolution of the racemate using  
CC dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.  
SQ Sequence 1442 BP; 385 A; 350 C; 370 G; 337 T;

|                       |  |               |           |              |
|-----------------------|--|---------------|-----------|--------------|
| Query Match           | 100.0%;  | Score 1442;   | DB 1;     | Length 1442; |
| Best Local Similarity | 100.0%;  | Pred. No. 0;  |           |              |
| Matches 1442;         | Conservative 0;  | Mismatches 0; | Indels 0; | Gaps 0;      |
| QY 1                  | ccccgggaactccatgtggccgtgatactctgctgagcaggatattgcatgattccagcggg | 60            |           |              |
| DB 1                  | ccccgggaactccatgtggccgtgatactctgctgagcaggatattgcatgattccagcggg | 60            |           |              |
| QY 61                 | cgcacagcgtctgctgagcaggatattgataaaaggcctggttagaagcgtgacccaacaac | 120           |           |              |
| DB 61                 | cgcacagcgtctgctgagcaggatattgataaaaggcctggttagaagcgtgacccaacaac | 120           |           |              |
| QY 121                | agctctctgatgatcttttaatgcgtcgtcatctgctctgtaactaaacgcgtataaatt   | 180           |           |              |
| DB 121                | agctctctgatgatcttttaatgcgtcgtcatctgctctgtaactaaacgcgtataaatt   | 180           |           |              |
| QY 181                | acgtggagaataacatataaataagtggttggaagaatccattatggccaacgcggtgttg  | 240           |           |              |
| DB 181                | acgtggagaataacatataaataagtggttggaagaatccattatggccaacgcggtgttg  | 240           |           |              |

```
Db 181 ACGTCGAGAAATACATATATCAAAATGGTTGGAAGAAATCCATTATGGCCAAACGCGGTGG 240
QY 241 tgccggggttaaccggtaacgcatcacctgacggaagaaatgcaaaagagtttccatta 300
Db 241 TGCCGGGGTAAACCGGTAAACGCATCACTGACGGAAGAAATGCAAAAGAGTTTCATTA 300
QY 301 caccattggcccttattccacacccgctctctgaccatcgaaacccggtgaccggattattgt 360
Db 301 CACCATTGGCCCTTATTCCACACCGCTCTGACCACTCGAACCCGCTGACCGGATTATTGT 360
QY 361 cgacactcgagatgcttttgaagtgctatcaattcgaacagagatatccgagccatt 420
Db 361 CGACACTCGAGATGCTTTTGAAAGTGCTATCAATTGCGAACAGGATATTCGAGCCAGTT 420
QY 421 gtaaaaaatgccctttctcaaccccaaaaaacggaccgatcatggtcaatggcggagaaa 480
Db 421 GCTAAAAATGCCCTTCTCAACCCCAAAAACGGACCGATCATGTGTCATATGGCGGAGAA 480
QY 481 aggtgatgtgctgctgtctatatgaatccatgttgcgccggcggttgatccctacgg 540
Db 481 AGGTGATGTGCTGCTGTCTATATCGAATCCATGTGTGCCCGCGGCTTGATCCCTACGG 540
QY 541 catctgcgccatgattccgcatcttggcgactgacccgggacccacctgacggccaatct 600
Db 541 CATCTGCCCATGATTCGCCATTTTGGCGGACTGACCCGGACCGACCTGACGGCCATGCT 600
QY 601 caatgatccgctgccagaaaaggctgcgatgattaaactcgacagtgaaaaggctactg 660
Db 601 CAATGATCCGCTGCCAGAAAAGGTGCGCATGATTAAATCGACAGTGAAGAGTCTACTG 660
QY 661 gagcaaacgcatacgtctccctataaaccataatgttgcccgccggttgatccctacga 720
Db 661 GAGCAAAACGCATAGCTCTCCCTATAAACCCCATATTGGCCACCTTGACGCTATCGCCAGA 720
QY 721 aattgactcaatcaattcaactgaagcagacaataccagcgcggaatatggatggtgcgga 780
Db 721 AATTGACTCAATCAATTCACTGACGCCAGACAATCAGCGCGGGAATATGGATGTGCCGGA 780
QY 781 tataggaccaggagatatacctatctgcggttaoatgctgcgctggagccgctgtttat 840
Db 781 TATAGGACACGAGGAGTATTACCTATCTGCCGGTACGTGCGCTGGAGCGCCCTGTTTAT 840
QY 841 tggtagcccatctgtcaggtgatgtgagatttgaggattggccgagcagtagatttgc 900
Db 841 TGGTAGTCCCATCTGTCTGAGGTGATGTTGAGATTGTCGGGACCGCAGTAGAGTTGC 900
QY 901 ctcaatcaccaccatcaaatgcattgatcaagaactggcagctttctctggccacgaat 960
Db 901 CTCAATCACCACCACCAATCAAGTCGATTGATCAAGAAGTGGCAGCTTCTGCGCCACGAAT 960
QY 961 ggagaatgccaaaatatattagatttgccagtgacgtcgcgtgcgtgagatgcgacgcg 1020
Db 961 GGAGAATGCCAAAATATTATGATTTATGGCAGTGCACGTCCGCTGGAGGATGGACGCG 1020
QY 1021 aattgcataatcgcaacttaattactgctggttagaagactttgcttcgaacaatggga 1080
Db 1021 AATTGCATATCGGCACCTTAATTACTGCTGGTAGAGACTTTTGGCTTCGAACAATGGGA 1080
QY 1081 tgcctacatgcttctgagtcgaatgcgcaaaagtgcgctgggcaacatggtcgaccacca 1140
Db 1081 TGCTTACATGCTTCTGAGTCAATCGGCAAGTGCAGCTGGGCAACATGTTGCGACCCCAA 1140
QY 1141 atacacgcttgccgcatgctgaacaaaaccctgttagtttagtagaataaactaacgg 1200
Db 1141 ATACACCGTTGGCGCATGTGAACAAAAACCTGTTAGTTTAGTAGAATAAATAAACCAG 1200
QY 1201 tgaacattaccggatgtagatcgggtaattgtaagtccaacacgcctatttttaa 1260
Db 1201 TGAACATTACCGGATGATCGGGTAAATGTAGTAAGTTCAACAATATCGCTAATTTTAA 1260
QY 1261 cagctaaagcaggatgcataatggggccagatacaccccaatattggtttactttacc 1320
Db 1261 CAGCTAAAGCAGGTGCATATGGGGCCAGATACCCCATCAATATATGTTTACTTTACTCC 1320
```

```
QY 1321 ttcaagcgagtgacggcgacacaaagagttgtcaacaatgacggagcaaccaggctatt 1380
Db 1321 TTCAGCGAGTGACGGCGGCACAGAGTTGTACAAATGGCGGAGCAACCCAGGCTATT 1380
QY 1381 gccaaaatataatcaaaatggcgcatcaacgacgagaccactcaatgcaattcatttgat 1440
Db 1381 GCCAAAATTAATCAAAATGGCGGCATCAACGCGAGACCACCTCAATGCAATTCATTGGAT 1440
QY 1441 cc 1442
Db 1441 CC 1442

RESULT 2
Q66950/C
ID Q66950 standard; cDNA; 3900 BP.
AC Q66950;
DE 15-FEB-1995 (first entry)
DE Cell death hid cDNA.
KW Cell death; head involution defective; hid gene; reaper gene; rpr;
KW apoptosis; transgenic animal; antisense; ss.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT cds 400..1632
FT /*tag= a
WO9416071-A.
PN 21-JUL-1994.
PF 14-JAN-1994; U00500.
PR 15-JAN-1993; US-004957.
PR 17-SEP-1993; US-123343.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Abrams JM, Grether ME, Steller H, White K;
DR WPI: 94-249218/30.
DR P-PSDB: R55791.
PT Isolated cell death genes from Drosophila - and novel assays for
PT apoptotic cell deaths and apoptotic and necrotic cell deaths,
PT based on selective staining by toluidine blue, acridine orange
PT and Nile blue
PS Disclosure: Fig. 6; 93pp; English.
CC The reaper (rpr) and head involution defective (hid) genes, mapping
CC to position 75C1,2 of D. melanogaster chromosome 3, exhibit
CC expression patterns related to the pattern of cell death during
CC Drosophila embryogenesis. Cell death genes, or antisense sequences,
CC can be used to reduce or abolish apoptosis, e.g. in transgenic
CC animals. An hid cDNA sequence is given in Q66950, and the
CC deduced sequence of the encoded HID protein in R55791.
SQ Sequence 3900 BP; 1288 A; 876 C; 736 G; 1000 T;

Query Match 2.7%; Score 39.2; DB 1; Length 3900;
Best Local Similarity 49.1%; Pred. No. 0.0077;
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
QY 320 acaccgcctctgaccatcgaaacccggtgaccgattattgtcgacactcgagatgtttt 379
Db 1458 ACCACCGAGCTGGCTGTCTGATGGCAGACTGGATTATTGCTGCTCGAGTGGCTATT 1399
QY 380 gaagggtgctaatcgaacagagatatccgagccagttgctaaaaatgccctttctc 439
Db 1398 GGAGCTCTTCTTCTTCGGGGGGGATGCGTCCATTGAACTCCTGCGAGTGGCTGTCG 1339
QY 440 aaccacaaaacgaccgatcatggttcaatggcgcgagaaagggtgatgtgctcgtc 499
Db 1338 GGCCGTAAAGTTTCTGCTAGCGATCGCAAACTCGTCCCAAGTGGCGTCATGATCGTCTG 1279
QY 500 tatatgaatccatgttgcgcgcgcgcgttga 531
Db 1278 GTACTCGCGCTCATCTCGCCCGCTCGAGGTCA 1247
```

```
RESULT 3
Q66948/C
```

ID Q66948 standard; DNA; 502 BP.  
AC Q66948;  
DE 15-FEB-1995 (first entry)  
DE Cell death hid gene.  
KW Cell death; head involution defective; hid gene; reaper gene; rpr;  
KW apoptosis; transgenic animal; antisense; ss.  
OS Drosophila melanogaster.  
FH Key Location/Qualifiers  
FT misc\_difference 93  
FT /\*tag= a  
FT /note= "Base n at position 93 is not identified  
FT in the specification"  
FN W09415071-A.  
PD 21-JUL-1994.  
PF 14-JAN-1994; U00500.  
PR 15-JAN-1993; US-004957.  
PR 17-SEP-1993; US-123343.  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
PI Abrams JM, Grether ME, Steller H, White K;  
DR WPI; 94-249218/30.  
DR P-PSDB; R55790.  
PT Isolated cell death genes from Drosophila - and novel assays for  
PT apoptotic cell deaths and apoptotic and necrotic cell deaths,  
PT based on selective staining by toluidine blue, acridine orange  
PT and Nile blue  
PS Disclosure; Fig.5; 93pp; English.  
CC The reaper (rpr) and head involution defective (hid) genes, mapping  
CC to position 75C1,2 of D. melanogaster chromosome 3, exhibit  
CC expression patterns related to the pattern of cell death during  
CC Drosophila embryogenesis. Cell death genes, or antisense sequences,  
CC can be used to reduce or abolish apoptosis, e.g. in transgenic  
CC animals. The genomic sequence of hid is given as 4 contiguous  
CC sequences in Q66946-49, which together encode HID protein.  
CC Sequence 502 BP; 141 A; 135 C; 135 G; 90 T;  
SQ

Query Match 2.6%; Score 37.6; DB 1; Length 502;  
Best Local Similarity 48.6%; Pred. No. 0.0072;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
Oy 320 acaccgctcctgaccatcgaaaccggtgaccggtattgttcgacacccgagatgcttt 379  
Db 333 ACCACCGACATGGCTGTCGGTATGCGACACTGGATTATTGCTGCTGCGAGTGGCTATT 274  
Oy 380 gaagtgctatcaatcggaacagatattccgagccagctggtgataaaatgccctttctc 439  
Db 273 GGAGCTCTCTCTTCCTCGGGGGGGATGCGTCCATTGAACCTCTCGACACGCTCCGGCC 214  
Oy 440 aaccacaaaacgaccgatcatgtcgaatggcgcggaaggtgatgtcgtcgtgctc 499  
Db 213 GGCCGTAAGTTGTGTAGCGATCGCAAACTCGTCCCAAGTGGCTCATGATCGCTCTG 154  
Oy 500 tatatcgaatccatgttgcggcgccgcttga 531  
Db 153 GTACTCGGCTCATCTCGCCCGTCGAGGTCA 122

RESULT 4  
X28300/C  
ID X28300 standard; DNA; 429 BP.  
AC X28300;  
DE 17-JUN-1999 (first entry)  
DE Human CYP3A4 gene exon 7, intron 7.  
KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;  
KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;  
KW genetic linkage detection; phenotypic variation; intron; ss.  
OS Homo sapiens.  
PN W09913106-A1.  
PD 18-MAR-1999.  
PF 02-SEP-1998; U18158.  
PR 10-SEP-1997; US-058612.  
PA (AXIS-) AXIS PHARM INC.  
PI Guida M, Lichter JB;

DR WPI; 99-215070/18.  
PT New isolated CYP3A4 polymorphic sequences  
PS Disclosure; Page 30; 40pp; English.  
CC This sequence represents an intron of the human CYP3A4 gene.  
CC The invention relates to a CYP3A4 sequence polymorphism,  
CC which is part of a non-naturally occurring chromosome. Nucleic acids  
CC comprising the CYP3A4 polymorphic sequences can be used to screen  
CC patients for altered metabolism for CYP3A4 substrates, potential  
CC drug-drug interactions, and adverse/side effects as well as diseases that  
CC result from environmental or occupational exposure to toxins. They can  
CC also be used to establish animal, cell culture and in vitro cell-free  
CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used  
CC for expression studies to determine the effect of promoter and/or intron  
CC sequence variations on mRNA expression and stability. The polymorphisms  
CC are also used as single nucleotide polymorphisms to detect genetic  
CC linkage to phenotypic variation in activity and expression of CYP3A4. The  
CC nucleic acids can also be used to generate genetically modified non-human  
CC animals or site specific gene modifications in cell lines.  
SQ Sequence 429 BP; 126 A; 95 C; 53 G; 153 T;

Query Match 2.4%; Score 35.2; DB 1; Length 429;  
Best Local Similarity 57.5%; Pred. No. 0.044;  
Matches 61; Conservative 1; Mismatches 44; Indels 0; Gaps 0;  
Oy 1164 acaaaaacctgttagtttagtaagaataacacggtgaacattaccggtgtagtc 1223  
Db 315 ACACATATCTTCAATGTACTACAAATCACTGAACCTGTATTTTAAAGTGAATTAC 256  
Oy 1224 ggggtgaagtgtgaatgcaacaaatcgctattttaacagctaaag 1269  
Db 255 ATGGTGATTATATCTCAATAAAGCAGTTATTTTAAAGAGAGMAAG 210

RESULT 5  
X00477  
ID X00477 standard; DNA; 5733 BP.  
AC X00477;  
DT 30-MAR-1999 (first entry)  
DE Arabidopsis thaliana clavatal encoding DNA.  
KW Arabidopsis thaliana; clavatal; CIV1; meristem phenotype; harvesting;  
KW apical shoot enlargement; floral meristem; resistance; ss.  
OS Arabidopsis thaliana.  
FH Key Location/Qualifiers  
FT CDS 2434..5470  
FT /\*tag= a  
FT /note= "contains introns"  
FT exon 2434..5038  
FT /\*tag= b  
FT intron 5039..5117  
FT /\*tag= c  
FT exon 5118..5470  
FT /\*tag= d  
FT /number= 2  
PN US5859338-A.  
PD 12-JAN-1999.  
PF 06-JUN-1995; 473553.  
PR 06-JUN-1995; US-473553.  
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
PI Clark SE, Meyerowitz EM, Williams RW;  
DR WPI; 99-120031/10.  
DR P-PSDB; W30607.  
PT Isolated nucleic acid encoding the plant clavatal protein and  
PT related vectors, plant cells and plants - used to regulate meristem  
PT phenotype, for increasing yield of leaves, flowers, fruits and seeds  
PS Claim 3; Fig 5; 45pp; English.  
CC The present sequence encodes clavatal (CIV1) isolated from Arabidopsis  
CC thaliana. The nucleotide sequence encoding CIV1, including its  
CC corresponding antisense sequences, are used to alter meristem phenotypes,  
CC in particular the enlargement of apical shoot or floral meristems, to  
CC provide an increased yield of leaves, flowers, fruits and seeds in a

CC very wide variety of plants. They may also be used to increase resistance  
 CC to lodging or wind damage, or to facilitate harvesting. Fragments of the  
 CC nucleotide sequence encoding CNV1 are also useful as probes to identify  
 CC related sequences in other plants.  
 SQ Sequence 5733 BP; 1732 A; 1144 C; 1094 G; 1763 T;

Query Match 2.4%; Score 34.2; DB 1; Length 5733;  
 Best Local Similarity 50.3%; Pred. No. 0.54;  
 Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1136 cccaatacaccttgccgcatgctgaacaaaaccctgttagtttagtaggaataaacta 1195  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 3709 CTCACGGCACGTGTCGCGGGGGCTTTCAATCATACCGTTAGTACGATTATCGAACTC 3768  
 QY 1196 accggtgaacattaccgcgatgtagtcgggtaagtgttaagtccaacaacatcgctatt 1255  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 3769 ACTGATAATTTCTTCGCGTGAACCTTCGGTAACGATGTCGGCGGATGTTCTCGATCAG 3828  
 QY 1256 tttaacagctaagcaggtgcataatggggccagatcacacccatcaat 1302  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 3829 ATTTACCTCTCTAACAACCTGGTTTCCGGCGGAGATTCCACCTGCCAT 3875

RESULT 6  
 X14578  
 ID X14578 standard; DNA; 745 BP.  
 AC X14578;  
 DT 31-MAR-1999 (first entry)  
 DE H. pylori GHPO 1662 gene.  
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease; ss.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FT CDS 37..720  
 FT /\*tag= a  
 PN W09843478-Al.  
 PD 08-OCT-1998.  
 PF 01-APR-1998; U06371.  
 PR 29-JUL-1997; US-902615.  
 PR 01-APR-1997; US-833457.  
 PR 24-JUN-1997; US-881227.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
 DR WPI; 98-342233/46.  
 DR P-PSDB; W98859.  
 DR New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 1; Page 1937-1938; 2054pp; English.  
 CC This sequence represents a polynucleotide of the invention. It was  
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.  
 CC The polypeptides can be used for preventing or treating Helicobacter  
 CC infections, and gastroduodenal diseases associated with these  
 CC infections, including acute, chronic, and atrophic gastritis, and peptic  
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used  
 CC for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 SQ Sequence 745 BP; 265 A; 102 C; 159 G; 219 T;

Query Match 2.3%; Score 33.2; DB 1; Length 745;  
 Best Local Similarity 48.9%; Pred. No. 0.32;  
 Matches 89; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 176 aaattacgtggagataacatatgaatgggttggaagaatccattatggccaaacgcggt 235  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 31 AAATTAATGGACAGAGAGGATGAATAATCAACAAATTAGTCATTATCAATCGGCC 90  
 QY 236 gttgggtccgggcgtaaacgcggttaacatcaacctgacggaagaataatgcaaaagagttt 295  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91 ATTCAGGTGGGGCAAGACCTCTTTGATCAACAGATTGAAGATGGCAAAAGCTTG 150

QY 296 cattacacacattggcccttatttccacacccgtcctgacacatgaacccgggtgacccgatt 355  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 151 GGCATTTCTATTAGCGTTCAITCTACCGATGAATATTTTCCAAACACAGATGAAGAGGT 210  
 QY 356 at 357  
 ||  
 Db 211 AT 212

RESULT 7  
 Q89701/c  
 ID Q89701 standard; DNA; 3201 BP.  
 AC Q89701;  
 DT 27-DEC-1995 (revised)  
 DT 31-OCT-1995 (first entry)  
 DE Alpha-1,4-glucan lyase gene.  
 KW Alpha-1,4-glucan lyase; enzyme; ss.  
 OS Morchella costata.  
 FH Key Location/Qualifiers  
 FT CDS 1..3201  
 FT /\*tag= a  
 PN W09510617-A.  
 PD 20-APR-1995.  
 PF 15-OCT-1994; E03398.  
 PR 15-OCT-1993; GB-021302.  
 PA (DANI-) DANISCO AS.  
 PI Boisen K, Christensen TMIE, Kragh KM, Marcussen J;  
 PI Yu S;  
 DR WPI; 95-161802/21.  
 DR P-PSDB; R72711.  
 DR Isolation of alpha-1,4-glucan lyase from fungus - and its  
 PT recombinant prodn from isolated DNA.  
 PS Claim 9; Pages 30-32; 51pp; English.  
 CC The enzyme alpha-1,4-glucan lyase(GL) was purified from Morchella  
 CC costata (Mc) by affinity chromatography on beta-cyclodextrin  
 CC Sepharose (RTM), ion exchange on Mono Q HR 5/5 (RTM) and gel  
 CC filtration on Sepharose (RTM) 12 columns. The purified enzyme  
 CC appeared colourless. The mol. mass was 110 kDa (SDS-PAGE). It showed  
 CC an isoelectric point of pI 5.4. The optimum pH range for the fungal  
 CC lyase catalysed reaction was between pH 5 and pH 7. The purified  
 CC fungal lyase degraded maltosaccharides. The highest activity was  
 CC with maltotetraose (activity 100%), then maltohexaose (97%),  
 CC maltotriose (56%) and maltose (2%). The lyase was digested with  
 CC endoproteinase Arg-C from Clostridium histolyticum or endoproteinase  
 CC Lys-C from Lysobacter enzymes. The resulting peptides were  
 CC sequenced (see FT). Amino acid sequences of three overlapping  
 CC tryptic peptides from GL (see R72713) were used to generate mixed  
 CC oligos which could be used as PCR primers for the amplification of  
 CC DNA isolated from Morchella. The primers are Q89703-4 and Q90305-6.  
 SQ Sequence 3201 BP; 903 A; 752 C; 770 G; 776 T;

Query Match 2.3%; Score 33.2; DB 1; Length 3201;  
 Best Local Similarity 50.0%; Pred. No. 0.83;  
 Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccgggcgttaaacgcgtaacgcacatcacgcggaagaatacgaaagagtttcattaca 302  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1028 CTGGGCTTCAACTTTGACCTGCCAACACACCTGTATAAGTCGTACAATTTCCGGGACC 969  
 QY 303 ccattggcccttattccacacccgctcctgacacccggtgaccggtgaccggtattgtcg 362  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 968 GTATCGCACTGATACCGTAACAAATCGATACCACCATCCTGGTTCCTCCAGCTTGATGTAG 909  
 QY 363 acactcgagatgcttttgagggtgctatcaattcgggaacaggatat 408  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 908 CTGAGTTGTTGCTTTTCCAAAGTCGTGGCAATTTGAGAGTAGTTAT 863

RESULT 8  
 Q87616/c  
 ID Q87616 standard; DNA; 3201 BP.



AC Q87616;  
DT 10-JAN-1996 (first entry)  
DE M.costata DNA encoding alpha-1,4-glucan lyase.  
KW Alpha-1,4-glucan lyase; 1,5-D-anhydrofructase; G.lemaleiformis;  
KW antioxidant; food additive; sweetener; ds.  
OS Morchella costata.  
FH key Location/Qualifiers  
FT 1. .3201  
FT cds  
FT /\*tag= a  
FT /product= alpha\_1,4-glucan\_lyase  
PD W09510616-A2.  
PN 20-APR-1995.  
PF 15-OCT-1994; E03397.  
PR 15-OCT-1993; GB-021302.  
PR 15-OCT-1993; GB-021301.  
PR 15-OCT-1993; GB-021303.  
PR 15-OCT-1993; GB-021305.  
PR 15-OCT-1993; GB-021304.  
PR (DANI-) DANISCO AS.  
PI Bojko M, Bojzen K, Christensen TMIE, Kragh KM, Marcussen J;  
PI Nielsen J, Yu S;  
DR WPI; 95-161801/21.  
DR P-PSDB; R70638.  
PT 1,5-D-anhydro:fructose prodn. from alpha-1,4-glucan - by treatment  
PT with pure alpha-1,4-glucan lyase, partic. useful as antioxidant and  
PT sweetener for foods and beverages  
PS Claim 12; Page 101c-102b; 166pp; English.  
CC An alpha-1,4-glucan lyase enzyme (preferably pullanase or isoamylase)  
CC is used in a new method for the production of 1,5-D-anhydrofructose.  
CC The enzyme is isolated from either a fungus (M.costata or M.vulgaris)  
CC or from fungally infected algae (G.lemaleiformis) or algae alone.  
CC 1,5-D-anhydrofructose is useful as an antioxidant and sweetener for  
CC foodstuffs and beverages. It is also useful as an intermediate for the  
CC antibiotic microthecin, an oxygen scavenger during polymerisation  
CC reactions and as a reducing agent in the synthesis of biodegradable  
CC plastics.  
SQ Sequence 3201 BP; 903 A; 752 C; 770 G; 776 T;

.Query Match 2.3%; Score 33.2; DB 1; Length 3201;  
Best Local Similarity 50.0%; Pred. No. 0.83;  
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccgggctaaacggtaacgcacacgcctcctgacgcaggaagaatgcacaaagagtttcattaca 302  
D 1028 CTGGGCTTCAACTTGAACGCTCAACAGACCTGTATAAAGTCGTACAAATTTCCGGGACC 969  
QY 303 ccattggcccttattccacacccgctcctgacacatgcaccccggtgacccgattattgtcg 362  
D 968 GTATCCGCACTGATACCGTAACAAATCGATACCACTACCTGTTCCACGCTTGATGTAG 909  
QY 363 acactcgagatgctttgaagtgctatcaattcgaacagagatat 408  
D 908 CCTGAGTTGGCTTTCCAAAGTCGATGGCAATTTGAGAGTAGTTAT 863

RESULT 9  
V84194/c  
ID V84194 standard; DNA; 3201 BP.  
AC V84194;  
DT 29-MAR-1999 (first entry)  
DE Morchella costata glucan lyase DNA.  
KW Glucan lyase; antioxidant; transgenic plant; stress tolerance;  
KW anhydrofructose; beverage; wine; foodstuff; ds.  
OS Morchella costata.  
PN W09850532-A2.  
PD 12-NOV-1998.  
PF 06-MAY-1998; IFB0708.  
PR 06-MAY-1997; GB-009161.  
PR (DANI-) DANISCO AS.  
PI Buchter-Larsen A, Marcussen I;  
DR WPI; 99-070094/06.  
DR P-PSDB; W88255.

PT Recombinant production of anti-oxidant compounds - by the production  
PT of an anhydrofructose from a glucan, used for improving plants for  
PT use in foodstuffs  
PS Claim 9; Page 45-46; 53pp; English.  
CC This nucleotide sequence codes for a glucan lyase (see W88255) of  
CC Morchella costata used in claimed processes of the invention. A  
CC claimed process of preparing a medium that contains an antioxidant  
CC and at least one other component involves preparing antioxidant  
CC from a glucan, in situ, by use of recombinant DNA techniques. Also  
CC claimed are: (1) use of anhydrofructose as an antioxidant for a  
CC medium, where the anhydrofructose is prepared in situ in the  
CC medium; (2) use of anhydrofructose for imparting or improving  
CC stress tolerance in a plant, where the anhydrofructose is prepared  
CC in situ in the plant; (3) use of glycane lyase for imparting or  
CC improving: (i) stress tolerance in plant, or (ii) transformation of  
CC grape, where the glucan lyase is prepared in situ; (4) use of a  
CC nucleotide sequence (NS) coding for a glucan lyase as a means of  
CC imparting or improving stress tolerance in a plant, where the NS is  
CC expressed in situ in the plant, and (5) use of a NS coding for a  
CC glucan lyase for imparting or improving the transformation of a  
CC grape, where the NS is expressed in situ in the grape. The  
CC antioxidant containing medium can be used as foodstuff or in the  
CC preparation of foodstuffs such as beverages, in particularly  
CC alcoholic beverages such as wine (claimed).  
SQ Sequence 3201 BP; 903 A; 752 C; 770 G; 776 T;

Query Match 2.3%; Score 33.2; DB 1; Length 3201;  
Best Local Similarity 50.0%; Pred. No. 0.83;  
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 ccgggctaaacggtaacgcacacgcctcctgacgcaggaagaatgcacaaagagtttcattaca 302  
D 1028 CTGGGCTTCAACTTGAACGCTCAACAGACCTGTATAAAGTCGTACAAATTTCCGGGACC 969  
QY 303 ccattggcccttattccacacccgctcctgacacatgcaccccggtgacccgattattgtcg 362  
D 968 GTATCCGCACTGATACCGTAACAAATCGATACCACTACCTGTTCCACGCTTGATGTAG 909  
QY 363 acactcgagatgctttgaagtgctatcaattcgaacagagatat 408  
D 908 CCTGAGTTGGCTTTCCAAAGTCGATGGCAATTTGAGAGTAGTTAT 863

RESULT 10  
V88779  
ID V88779 standard; cDNA; 383 BP.  
AC V88779;  
DT 12-FEB-1999 (first entry)  
DE EST clone H0266.  
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
OS Homo sapiens.  
PN W09845437-A2.  
PD 15-OCT-1998.  
PF 10-APR-1998; U06956.  
PR 10-APR-1997; US-837312.  
PR (GEMV) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI; 99-070078/06.  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries  
PS Claim 1; Page 509-510; 641pp; English.  
CC The present sequence represents an expressed sequence tag (EST), and is  
CC a polynucleotide of the invention. The polynucleotides of the invention  
CC are all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and



```

FT /label= AP1
FT 113. .131
FT /*tag= d
FT /label= NF-1
FT 154. .163
FT /*tag= e
FT /label= AP4
FT 171. .180
FT /*tag= f
FT /label= AP4
FT 201. .212
FT /*tag= g
FT /label= HNF-3
FT 284. .300
FT /*tag= h
FT /label= NF-1
FT 292. .300
FT /*tag= i
FT /label= AP4
FT 307. .316
FT /*tag= j
FT /label= AP4
FT 389. .400
FT /*tag= k
FT /label= AP4
FT 421. .450
FT /*tag= l
FT /label= HNF-4alpha
FT 459. .470
FT /*tag= m
FT /label= HNF-3
FT 479. .490
FT /*tag= n
FT /label= C/EBP
FT 492. .514
FT /*tag= o
FT /label= HNF-3
FT 516. .537
FT /*tag= p
FT /label= AP1
FT 715. .6215
FT /*tag= q
FT /note= "contains introns"
FT 490. .1040
FT /*tag= r
FT /number= 1
FT 1041. .1756
FT /*tag= s
FT /number= 1
FT misc_difference 1352
FT /*tag= t
FT /note= "n at position 1352 represents an intronic
FT sequence of 9 kb"
FT 1757. .1956
FT /*tag= u
FT /number= 2
FT 1957. .2588
FT /*tag= v
FT /number= 2
FT misc_difference 2205
FT /*tag= w
FT /note= "n at position 2205 represents an intronic
FT sequence of 3.8 kb"
FT 2589. .2752
FT /*tag= x
FT /number= 3
FT 2753. .3210
FT /*tag= y
FT /number= 3
FT 3211. .3452
FT /*tag= z
FT /number= 4
FT 3453. .3739
FT intron

```

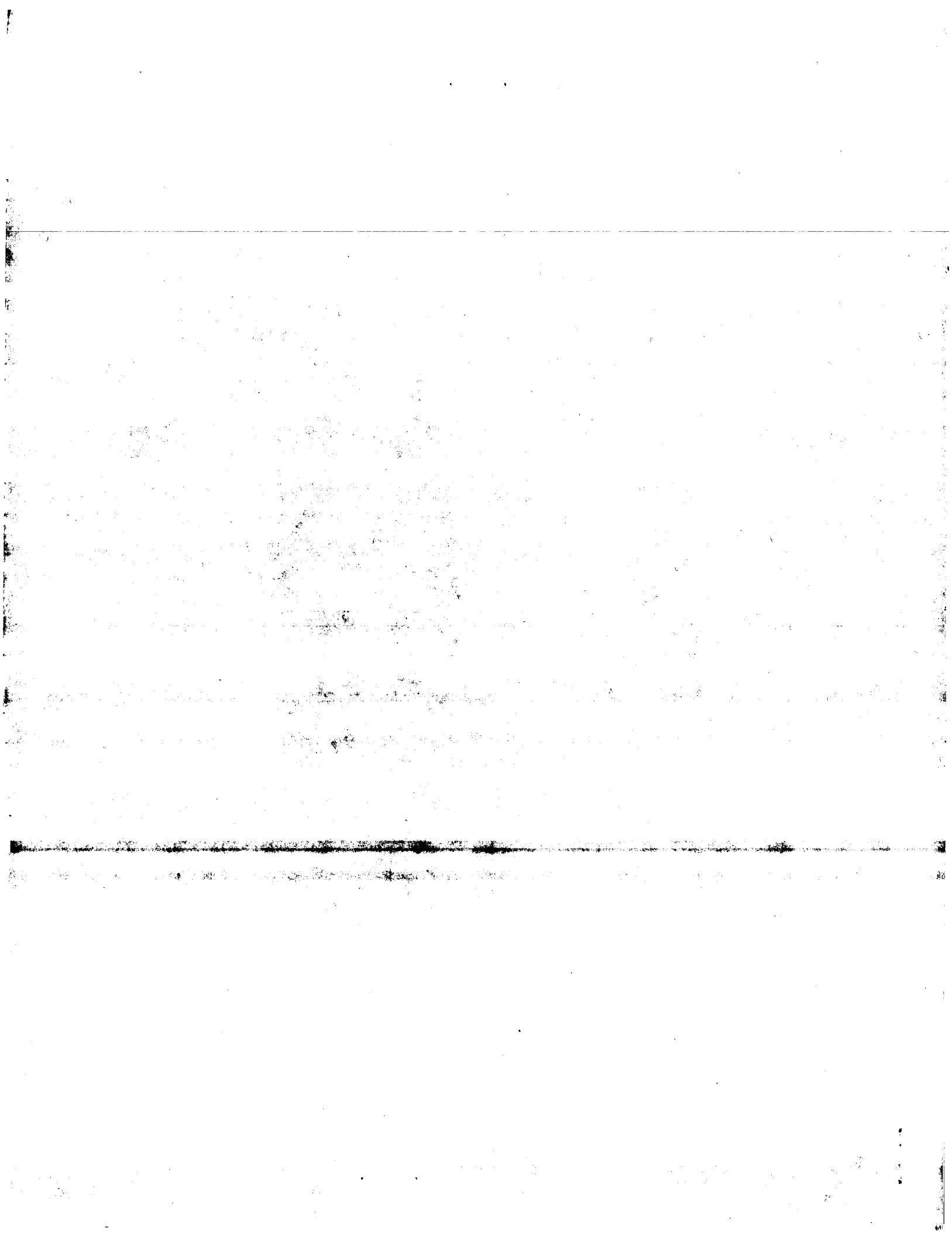
```

FT misc_difference 3564
FT /*tag= aa
FT /number= 4
FT /*tag= ab
FT /note= "n at position 3564 represents an intronic
FT sequence of 2 kb"
FT 3740. .3901
FT /*tag= ac
FT /number= 5
FT 3902. .4018
FT /*tag= ad
FT /number= 5
FT 4019. .4220
FT /*tag= ae
FT /number= 6
FT 4221. .4671
FT /*tag= af
FT /number= 6
FT misc_difference 4339
FT /*tag= ag
FT /note= "n at position 4339 represents an intronic
FT sequence of 0.8 kb"
FT 4672. .4863
FT /*tag= ah
FT /number= 7
FT 4864. .5326
FT /*tag= ai
FT /number= 7
FT misc_difference 5177
FT /*tag= aj
FT /note= "n at position 5177 represents an intronic
FT sequence of 1.5 kb"
FT 5327. .5448
FT /*tag= ak
FT /number= 8
FT 5449. .5541
FT /*tag= al
FT /number= 8
FT 5542. .5686
FT /*tag= am
FT /number= 9
FT 5687. .6087
FT /*tag= an
FT /number= 9
FT misc_difference 5853
FT /*tag= ao
FT /note= "n at position 5853 represents an intronic
FT sequence of 1.5 kb"
FT 6088. .6260
FT /*tag= ap
FT /number= 10
FT WO9811254-A1.
FT 19-MAR-1998.
FT 10-SEP-1997; U16037.
FT 30-OCT-1996; US-029679.
FT 10-SEP-1996; US-025719.
FT 02-OCT-1996; US-028056.
FT (ARCH-) ARCH DEV CORP.
FT Bell GI, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S,
FT Oda N, Yamagata K;
FT WPI; 98-271667/24.
FT DR P-PSDB; W71580.
FT Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and
FT 1-beta - useful for detecting susceptibility for non-insulin
FT dependent diabetes, especially maturity-onset diabetes of the young
FT PS Claim 95; Fig 26A-M; 363pp; English.
FT CC This is a partial nucleotide sequence of the gene (TCF2) encoding
FT human hepatocyte nuclear factor 1 beta (HNF-1 beta, see W71580), a
FT CC homeodomain-containing transcription factor. Mutations in this
FT CC gene are indicative of a propensity to diabetes mellitus. The
FT CC invention concerns the identification of genes responsible for
FT non-insulin dependent diabetes mellitus (NIDDM) for use in
FT CC diagnostics and therapeutics. It demonstrates that the MODY3 locus

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:16:07 ; Search time 52.2 Seconds  
(without alignments)  
3799.914 Million cell updates/sec

Title: US-09-214-679-1

Perfect score: 1442

Sequence: 1 cccgggaactccattgtgcc.....aatgcaattcattgtgatcc 1442

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 6877915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/PCrUS\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/2/ina/PCrUS\_COMB1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| C 1        | 37.6  | 2.6         | 503    | 1     | US-08-123-343A-15 |
| C 2        | 37.6  | 2.6         | 3900   | 1     | US-08-123-343A-6  |
| C 3        | 34.2  | 2.4         | 5733   | 3     | US-08-473-553A-1  |
| C 4        | 33.2  | 2.3         | 3201   | 3     | US-08-633-770A-3  |
| C 5        | 33.2  | 2.3         | 4726   | 3     | US-08-633-770A-11 |
| C 6        | 32.8  | 2.3         | 7218   | 1     | US-08-232-463-14  |
| C 7        | 32.4  | 2.2         | 9551   | 1     | US-08-056-200-93  |
| C 8        | 32.4  | 2.2         | 9551   | 4     | US-08-800-644-93  |
| C 9        | 32.2  | 2.2         | 1226   | 1     | US-08-374-686-4   |
| C 10       | 32.2  | 2.2         | 1349   | 1     | US-08-374-686-1   |
| C 11       | 31.2  | 2.2         | 2167   | 4     | US-08-743-637B-16 |
| C 12       | 31.2  | 2.2         | 2167   | 5     | US-08-526-840B-16 |
| C 13       | 30.8  | 2.1         | 1187   | 6     | PCr-US95-13536-2  |
| C 14       | 30.8  | 2.1         | 1188   | 6     | PCr-US95-13536-1  |
| C 15       | 30.2  | 2.1         | 10095  | 5     | US-08-822-586-45  |
| C 16       | 30    | 2.1         | 2665   | 5     | US-09-040-005-1   |
| C 17       | 29.4  | 2.0         | 1888   | 2     | US-08-485-107-1   |
| C 18       | 29.2  | 2.0         | 46899  | 2     | US-08-471-119A-1  |
| C 19       | 29    | 2.0         | 300    | 1     | US-08-083-946-6   |
| C 20       | 29    | 2.0         | 300    | 5     | US-08-452-915-6   |
| C 21       | 29    | 2.0         | 850    | 1     | US-07-611-528A-1  |
| C 22       | 29    | 2.0         | 850    | 1     | US-08-083-946-1   |
| C 23       | 29    | 2.0         | 850    | 5     | US-08-452-915-1   |
| C 24       | 28.6  | 2.0         | 1212   | 4     | US-09-092-770-18  |
| C 25       | 28.4  | 2.0         | 2380   | 7     | 5268463-1         |
| C 26       | 28.4  | 2.0         | 2438   | 7     | 5432081-1         |

Query Match 2.6%; Score 37.6; DB 1; Length 503;

ALIGNMENTS

RESULT 1

US-08-123-343A-15/c

; Sequence 15, Application US/08123343A

; Patent No. 5593879

; GENERAL INFORMATION:

; APPLICANT: Steller, Hermann

; APPLICANT: Abrams, John M.

; APPLICANT: Grether, Megan E.

; APPLICANT: White, Kristin

; TITLE OF INVENTION: Cell Death Genes of Drosophila

; TITLE OF INVENTION: Melanogaster and Vertebrate Analogs

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: MA

; COUNTRY: US

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/123,343A

; FILING DATE: 17-SEP-1993

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/004,957

; FILING DATE: 15-JAN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Granahan, Patricia

; REGISTRATION NUMBER: 32,227

; REFERENCE/DOCKET NUMBER: MIT-5907A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-861-6240

; TELEFAX: 617-861-9540

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 503 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; US-08-123-343A-15





Qy 363 acactcgagatgcttttgaagggtgctatcaattcgaacaggatat 408  
 Db 908 CCTGAGTTGGTCTTTCCAAAGTCGATGGCAATTTGACAGTAGTTAT 863

RESULT 5  
 US-08-633-770A-11/c  
 ; Sequence 11, Application US/08633770A  
 ; Patent No. 5908760  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bojsen, Kirsten  
 ; APPLICANT: Yu, Shukun  
 ; APPLICANT: Kragh, Karsten  
 ; APPLICANT: Christensen, Tove  
 ; APPLICANT: Marcussen, Jan  
 ; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
 ; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive 16th Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.

```

; DATE: 02-08-90
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Patatin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: PCT/EP94/03398  
 FILING DATE: OCT-15-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Altman, Daniel E  
 REGISTRATION NUMBER: 34,115  
 REFERENCE/DOCKET NUMBER: DYO06.001APC  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 714-760-0404  
 TELEFAX: 714-760-9502  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4726 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-633-770A-11

|    | Query Match           | 2.3%   | Score 33.2;    | DB 3;     | Length 4736; |
|----|-----------------------|--|----------------|-----------|--------------|
|    | Best Local Similarity | 50.0%;   | Prod. No. 1.1; |           |              |
|    | Matches 83;           | Conservative 0;  | Mismatches 83; | Indels 0; | Gaps 0;      |
| Qy | 243                   | ccgggcgtaaaccggtaacgcacatcacctgcgcggaagaataatgcacaaagaggtttcattaca | 302            |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |
| Db | 1741                  | CTGGGCTTCAACTTTGAACGPTCCAACAGACTGTATAAAGTCGTACAATTTCCGGGACC        | 1682           |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |
| Qy | 303                   | ccattggcccttattccacacccgctcctgcacatcgaaaccgggtgaccgggattattgtcgcg  | 362            |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |
| Db | 1681                  | GTATCCGCACTGATACCGTACATCGATACCACTACCTGGTTCACAGCTTGATGTAG           | 1622           |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |
| Qy | 363                   | acactcgagatgcgttttgaagggtgctatcaattcgggaacaggatat                  | 408            |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |
| Db | 1621                  | CCTGAGTTGGTCTTTCCAAACTCCGATGGCAATTTTGAGAGTAGTAT                    | 1576           |           |              |
|    |                       |  |                |           |              |
|    |                       |  |                |           |              |

9





```
Db 508 GATGGAGCTATATTTTA 492
|||||
RESULT 10
US-08-374-686-1/c
; Sequence 1, Application US/08374686
; Patent No. 5616474
; GENERAL INFORMATION:
; APPLICANT: Bolotin, Monique
; APPLICANT: Menart, Sandrine
; TITLE OF INVENTION: K. lactis Transaldolase Gene Promoter
; TITLE OF INVENTION: and Use Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,686
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00771
; FILING DATE: 28-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/09432
; FILING DATE: 30-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92048-US
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1297..1347
US-08-374-686-1

Query Match 2.2%; Score 32.2; DB 1; Length 1349;
Best Local Similarity 63.6%; Pred. No. 1.1;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1160 ctgaacaaaacctgttagtttagtagaataactaacccggtgaacattaccggatgta 1219
|||||
Db 640 CGGATGGACACACCGGGTAGTTTATGCGGGATGATCTTCCAGTGATCTTCCCTGGTGGTA 581
|||||
QY 1220 gatcggggtaatgtgta 1236
|||||
Db 580 GATGGAGCTATATTTTA 564
|||||

RESULT 11
US-08-743-637B-16
; Sequence 16, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
US-08-743-637B-16

Query Match 2.2%; Score 31.2; DB 4; Length 2167;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1303 attgggttactttactcttcaggagtgacggcgacaaagttgtcacaaatggcgc 1362
|||||
Db 162 ATCGGCTGCTGTCTCCGAAACCGCGTCACCGCGGATATCGAGCGCTCGCACGCGTAT 221
|||||
QY 1363 ggagcaaccaggctattgccgaaataatcaaaatggcggaatcaacggcagaccactc 1422
|||||
Db 222 GCGCATTCGTCGCGGTGAGCAACTGAACCGCGAGGCGGCGTCCGCCGATC 281
|||||
QY 1423 aatgcattcatttgatcc 1442
|||||
Db 282 GAAACGCTGCCAGGACCC 301
|||||

RESULT 12
US-08-526-840B-16
; Sequence 16, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
```

```
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2167 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; US-08-526-840B-16

Query Match 2.2%; Score 31.2; DB 5; Length 2167;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1303 attggtttacttactccttcaggagtgacggcgacgaagagttgtcacaaatggcgc 1362
DB 162 ATCGGCTGCTGTCTCCGAACCGCGTCACCGCGATATCGAGCGCTCGCAGCGGTAT 221
QY 1363 gagacacccaggctattgccgaataatcaaaatggcgcatcaacggcaccactc 1422
DB 222 GCGCATGTCGTCGGTCGAGCAACTGAACCGCGAGGCGGCTCGCGCGTCCCGCATC 281
QY 1423 aatgcaattcatttgatcc 1442
DB 282 GAAACGCTGTCCAGGACCC 301

RESULT 13
; Sequence 2, Application PC/TUS9513536
; GENERAL INFORMATION:
; APPLICANT: SKOPEK, THOMAS R
; TITLE OF INVENTION: SYNTHESIS OF METHYLASE RESISTANT GENES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELLIOTT KORSEN
; STREET: 126 E. Lincoln Avenue, P.O. Box 2000
```

```
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13536
; FILING DATE: 13-OCT-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSEN, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19257Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-5493
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; PCT-US95-13536-2

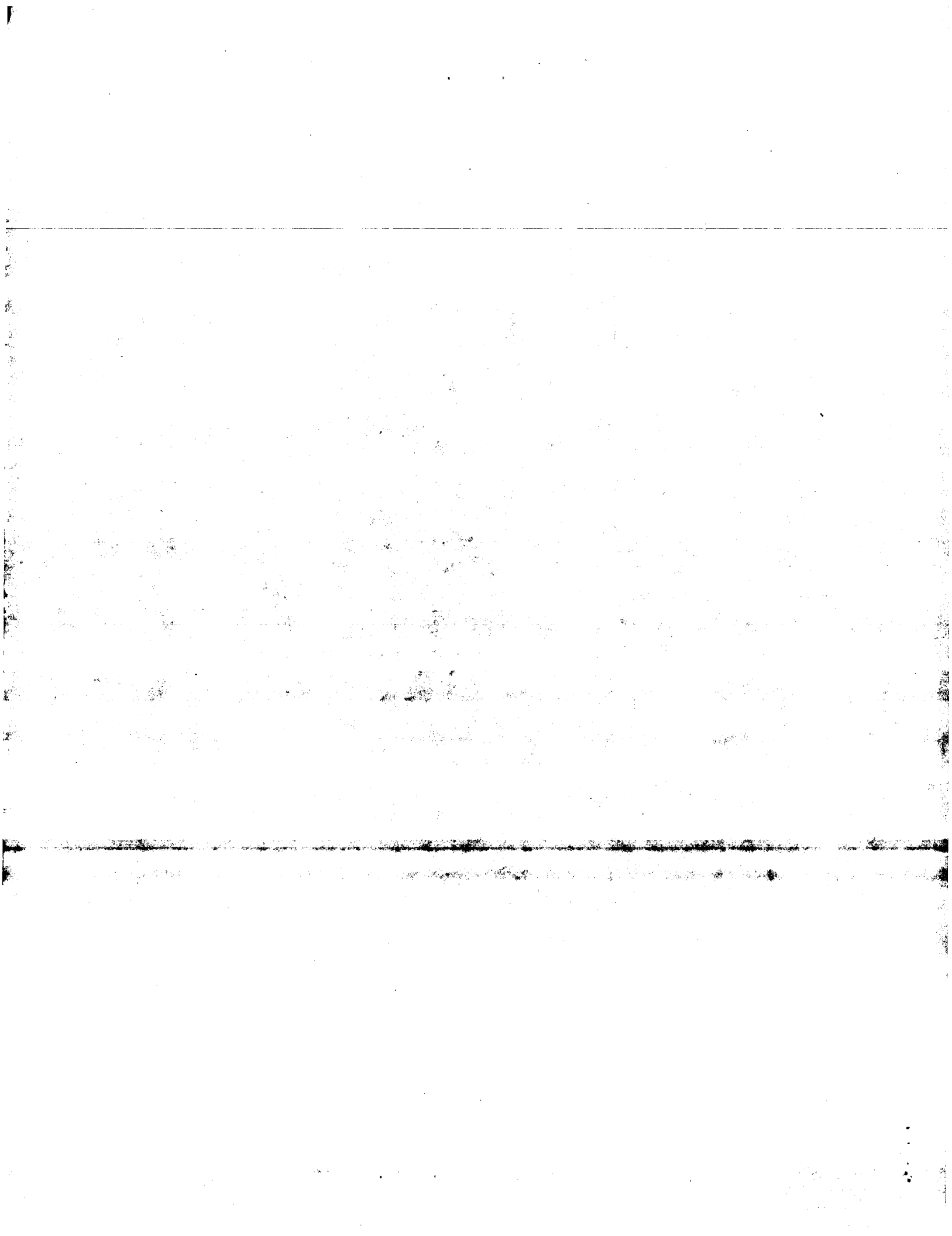
Query Match 2.1%; Score 30.8; DB 6; Length 1187;
Best Local Similarity 55.7%; Pred. No. 2.9;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 822 ctggagggccctgtttattggtagtcccacatgcttgcaggggtgagatttgcg 881
DB 868 CTGAATCTGGCTGCGTGTGGTGCAGATATCTGTGGTGGGCTATGATGATCTGAAG 927
QY 882 ggaccgcagtagatttgcctcaatcacaccaccatcaaaagtcgattt 927
DB 928 ACAGCTCATGTTATATCCACCAAGCACCACCATCAACACAGGATTT 973

RESULT 14
; Sequence 1, Application PC/TUS9513536
; GENERAL INFORMATION:
; APPLICANT: SKOPEK, THOMAS R
; TITLE OF INVENTION: SYNTHESIS OF METHYLASE RESISTANT GENES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELLIOTT KORSEN
; STREET: 126 E. Lincoln Avenue, P.O. Box 2000
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13536
; FILING DATE: 13-OCT-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSEN, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19257Y
```









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2000, 20:02:07 ; Search time 588.37 Seconds  
(without alignments)  
10807.778 Million cell updates/sec

Title: US-09-214-679-1  
Perfect score: 1442  
Sequence: 1 ccgggaatccatgtgcc.....aatgcaattcatttgatcc 1442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues  
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: gb\_gss7:\*  
104: gb\_gss8:\*  
105: gb\_gss9:\*  
106: em\_gss5:\*  
107: em\_gss6:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: em\_gss9:\*  
111: em\_gss10:\*  
112: em\_gss11:\*  
113: gb\_gss10:\*  
114: gb\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*



**JOURNAL**

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

**COMMENT**

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in plecter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

IMAGE:232927 5', mRNA sequence.
H75593.1 GI:1049521
EST.
LOCUS
DEFINITION
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 421)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On Mar 12, 1999 this sequence version replaced gi:4057371.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1012
High quality sequence stops: 366
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1012 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 366.
Location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="GDB:3786104"
/db_xref="taxon:9606"
/clone="IMAGE:232927"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/Note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAGAGTAATTAAGATCTTTTATTTTATTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
89 a 104 c 107 g 113 t 8 others
BASE COUNT
ORIGIN
Query Match 2.4%; Score 35; DB 86; Length 421;
Best Local Similarity 51.6%; Pred. No. 4.1; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 75;
QY 1044 actggctgtagaagactttggcttcgacaatggagctcacatgcttctgagtcatt 1103
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 ACTGGGAGGAGGATGACATTTGGCGCTCGGAGGAGATGATCTCAACGCGAGCTCGAGACAAT 183
QY 1104 gcggcaaaagtcggctgggcaacatggtcgaccccaataacacgcttggcgcatgctga 1163
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 CCCACAGGAGGATTCCTGGGGCGGAATGGCGGTGCACAGTTACACCAGGTTGGGGTGTAACC 123
QY 1164 acaaaaacctgttagtttagtaggaataaactaac 1198
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 ACTATACCTTTATATAGAGCTGATCTTTCTACCC 88

```



```

QY 1242 aaacaatcgctatttttaacagctaaa 1268
Db 543 AATATATCCAAAATATGATCATTTCAA 569

RESULT 8
A0610392/c
LOCUS A0610392 396 bp DNA GSS 15-JUN-1999
DEFINITION HS_5101_B2_A01_SP6E RPCI-11 Human Male BAC Library Homo
sapiens genomic clone Plate=677 Col=2 Row=B, genomic survey
sequence.
ACCESSION A0610392

VERSION A0610392.1 GI:5071668
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 396)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT On Feb 19, 1999 this sequence version replaced gi:4146198.
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 677 row: B column: 2
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 396.
FEATURES
source
1..396
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=677 Col=2 Row=B"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 128 a 113 c 66 g 89 t
ORIGIN

Query Match 2.4%; Score 34.8; DB 105; Length 396;
Best Local Similarity 47.3%; Pred. No. 4.7;
Matches 105; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 827 ggcgcgtgttatgtggtgccatgctgttcagggtgatgtgagattgcgggacc 886
Db 358 GGC CGCGGGTTGTGTGTGTGCCATGTCATTTGGTTAGGCCACGCTTGGGGTTT 299

QY 887 gcagtagatttgctcaatcacaccatcaagaagtcgattatcaagaactggcagctt 946
Db 298 ACAGAGCCGTGGCCACACAGGTCGGAAGGAATGGTGGAGTTTCTTTGGGGTTTAGGGTT 239

```

```

QY 947 tcttgccacgaatggagaatgccgaaataattatgattggcagtcgacgtcgctg 1006
Db 238 TGGGTCTACATTAGTGAACCTTTGGATAATTTTATAGGGTAGAGGGTCCGCTCGCTT 179

QY 1007 gaggatgcgacgcgaattgcataatgcgacttaatttctgg 1048
Db 178 TACCAAGGTAGCAGTCTTTCACAGTGCCTATTTATTAATAGG 137

RESULT 9
A0555596/c
LOCUS A0555596 422 bp DNA GSS 29-MAY-1999
DEFINITION HS_5229_A1_A09_SP6E RPCI-11 Human Male BAC Library Homo
sapiens genomic clone Plate=805 Col=17 Row=A, genomic survey
sequence.
ACCESSION A0555596

VERSION A0555596.1 GI:4915603
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 422)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT On Dec 15, 1999 this sequence version replaced gi:4215100.
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 805 row: A column: 17
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 422.
FEATURES
source
1..422
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=805 Col=17 Row=A"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 143 a 76 c 58 g 144 t
ORIGIN

```

```

Query Match 2.4%; Score 34.8; DB 104; Length 422;
Best Local Similarity 49.5%; Pred. No. 4.8;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 587 ctgacggccatgctcaatgattccgtgccagaaagggtgcgcatgattaaactcgacgt 646
Db 421 CAGTAGTAAAGCGTAATGCTCAATATCCTTAATTGTAGGTAGATATAAACTAAAGTGT 362

```

```

QY 647 gaaaagtgctactggagcaaacgccatacagcttcctccatataaaaccccatattggcaaccttg 706
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GTATAAGTCTATCAGACGACATGAGTTTGTATAGATCTGTTAAAGACAGACATAATAGCA 302
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 707 agcgtatcgccagaaattgactcaactcaattcaactacgacccagacacacacacgcggaat 766
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AGTTGATGACGTAATTTGTTTACTTTACCACTTATTACTTTTAAAGCATACCGTGTAT 242
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 767 at 768
|||
Db 241 AT 240

RESULT 10
AI743670/c
LOCUS
DEFINITION
wq4le06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2367682.3' similar to gb:Z23090 HEAT SHOCK 27 KD PROTEIN
(HUMAN); contains MER22.t1 YAK1 repetitive element ;, mRNA sequence.
ACCESSION
AI743670
VERSION
AI743670.1 GI:5111958
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 475)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 823 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 437.

FEATURES
source
1. .475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2367682"
/lab_host="Soares_NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-728407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 89 a 118 c 166 g 102 t
ORIGIN

Query Match 2.4%; Score 34.8; DB 43; Length 475;
Best Local Similarity 60.6%; Pred. No. 5;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 507 aatccatgttgcccggttgatccctacgcatctcgccatgattccgcatttg 566
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 AATACAGCTGCCCGCGGTGTGGACCCCAAGTTTCCTCTCCCTTTCCTCGAGG 262
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 567 gcggactgaccgggaccgacactgacggccatgct 600
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 GCACACTGACCGTGAAGGCCCCCATGCCCAAGCT 228
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AI346857/c
LOCUS
DEFINITION
qp53e12.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1926766.3'
similar to gb:Z23090 HEAT SHOCK 27 KD PROTEIN (HUMAN);, mRNA
sequence.
ACCESSION
AI346857
VERSION
AI346857.1 GI:4084063
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 477)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Apr 7, 1998 this sequence version replaced gi:3036231.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Seq primer: -400P from Gibco
High quality sequence stop: 1.

FEATURES
source
1. .477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1926766"
/lab_host="NCI_CGAP_Co8"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 93 a 118 c 152 g 114 t
ORIGIN

Query Match 2.4%; Score 34.8; DB 38; Length 477;
Best Local Similarity 60.6%; Pred. No. 5;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 507 aatccatgttgcccggttgatccctacgcatctcgccatgattccgcatttg 566
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 AATACAGCTGCCCGCGGTGTGGACCCCAAGTTTCCTCTCCCTGTCCTCATGAGG 258
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 567 gcggactgaccgggaccgacactgacggccatgct 600
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 GCACACTGACCGTGGAGGCCCCCATGCCCAAGCT 224
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 12
AI289776/c
LOCUS
DEFINITION
  AI289776 502 bp mRNA EST 21-DEC-1998
  qw12a01.x1 NCI-CGAP.Ut3 Homo sapiens cDNA clone IMAGE:1990824 3'
  similar to gb:Z23090 HEAT SHOCK 27 KD PROTEIN (HUMAN);, mRNA
  sequence.
ACCESSION
AI289776
VERSION
AI289776.1 GI:3931617
KEYWORDS
  human.
SOURCE
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 502)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Jun 22, 1998 this sequence version replaced gi:3247021.
  Contact: Robert Strausberg, Ph.D.
  Tel: (301) 496-1550
  Email: Robert.Strausberg@nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Insert Length: 884 Std Error: 0.00
  Seq primer: -400P from Gibco
  High quality sequence stop: 196.
  Location/Qualifiers
    1..502
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:1990824"
    /clone_lib="NCI-CGAP-Ut3"
    /tissue_type="poorly-differentiated endometrial
    adenocarcinoma, 2 pooled tumors"
    /lab_host="DH10B"
    /note="organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
    Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
    Average insert size 1.45 kb. Life Technologies catalog #:
    11541-018"
  BASE COUNT      91 a 115 c 155 g 141 t
  ORIGIN

  Query Match      2.4%; Score 34.8; DB 37; Length 502;
  Best Local Similarity 60.6%; Pred. No. 5.1;
  Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

  QY 507 aatccatgttgcgcgcgttgatccctacgcgcacatcgcgcacatgattcgcaatttg 566
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 350 AATACACGCTGCCCGGCTGGGACCCACCCCAAGTTTCCTCTCTCTCTCTCTCTCTGAGG 291

  QY 567 gcgagtgaccggaccgacctgcgcgcacgt 600
      || ||||| || || || || || || || || || || || || || || || || || ||
  DB 290 GCACACTGACCGTGAAGGCCCGCCATGCCCAAGCT 257

  RESULT 13
  A0482699
  LOCUS
  DEFINITION
    A0482699 511 bp DNA GSS 24-APR-1999
    RPCI-11-240H11.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-
    240H11, genomic survey sequence.
  ACCESSION
    A0482699
  VERSION
    A0482699.1 GI:4670103
  KEYWORDS
    GSS.
  SOURCE
    ORGANISM
      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 562)
    Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
    Venter,J.C.
    Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
    Map Building
    Unpublished (1997)
    On Dec 15, 1999 this sequence version replaced gi:4212149.
    Other GSS: RPCI-11-240H11.TV
    Contact: Shaying Zhao, William Nierman, Mark Adams
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: hbe@tigr.org
    Clones are derived from the human BAC library RPCI-11. For BAC
    library availability, please contact Pieter de Jong
    (pieter@dejong.med.buffalo.edu). Clones may be purchased from
    BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
    Research Genet cs (info@resgen.com). BAC end search page:
    http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
    Seq primer: SP6
    Class: BAC ends.
  FEATURES
    source
    location
    1..511
    /organism="Homo sapiens"
    /db_xref="GDB:7591954"
    /db_xref="taxon:9606"
    /clone="RPCI-11-240H11"
    /clone_lib="RPCI-11"
    /sex="Male"
    /cell_type="Lymphocytes"
    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
    RPCI11 Human Male BAC Library"
  BASE COUNT      181 a 79 c 89 g 162 t
  ORIGIN

  Query Match      2.4%; Score 34.6; DB 103; Length 511;
  Best Local Similarity 51.6%; Pred. No. 5.9;
  Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

  QY 1138 caatcacaccttggcgcgatgctgaacaaacctgttagtttagtaggaataactaac 1197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 82 CACATTCACAGGAACTGGTGTAGAGCCACACATGTGTTTGGAGGGCATATATCAA 141

  QY 1198 cgggtgaacattaccgcggtatgatcggggtaagtgttaagtccaacaacatcgctatttt 1257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 142 CCCCAACAGACGCGAGGTTATAGAGAACTTTGAATCCCATGTTAAATACCACTATTTT 201

  QY 1258 taacagctaaagcagggtgcacatcgggccagat 1290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 202 TCATACTAAGGGCAGTGGCTAGACAGGAAAGAT 234

  RESULT 14
  A0478968/c
  LOCUS
  DEFINITION
    A0478968 562 bp DNA GSS 23-APR-1999
    RPCI-11-269K12.TV RPCI-11 Homo sapiens genomic clone RPCI-11-
    269K12, genomic survey sequence.
  ACCESSION
    A0478968
  VERSION
    A0478968.1 GI:4661087
  KEYWORDS
    GSS.
  SOURCE
    ORGANISM
      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 562)
    Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
    Venter,J.C.
    Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
    Map Building
    Unpublished (1997)
    On Dec 15, 1999 this sequence version replaced gi:4212149.
    Other GSS: RPCI-11-240H11.TV
    Contact: Shaying Zhao, William Nierman, Mark Adams
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: hbe@tigr.org
    Clones are derived from the human BAC library RPCI-11. For BAC
    library availability, please contact Pieter de Jong
    (pieter@dejong.med.buffalo.edu). Clones may be purchased from
    BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
    Research Genet cs (info@resgen.com). BAC end search page:
    http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
    Seq primer: SP6
    Class: BAC ends.
  FEATURES
    source
    location
    1..511
    /organism="Homo sapiens"
    /db_xref="GDB:7591954"
    /db_xref="taxon:9606"
    /clone="RPCI-11-240H11"
    /clone_lib="RPCI-11"
    /sex="Male"
    /cell_type="Lymphocytes"
    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
    RPCI11 Human Male BAC Library"
  BASE COUNT      181 a 79 c 89 g 162 t
  ORIGIN

  Query Match      2.4%; Score 34.6; DB 103; Length 511;
  Best Local Similarity 51.6%; Pred. No. 5.9;
  Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

  QY 1138 caatcacaccttggcgcgatgctgaacaaacctgttagtttagtaggaataactaac 1197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 82 CACATTCACAGGAACTGGTGTAGAGCCACACATGTGTTTGGAGGGCATATATCAA 141

  QY 1198 cgggtgaacattaccgcggtatgatcggggtaagtgttaagtccaacaacatcgctatttt 1257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 142 CCCCAACAGACGCGAGGTTATAGAGAACTTTGAATCCCATGTTAAATACCACTATTTT 201

  QY 1258 taacagctaaagcagggtgcacatcgggccagat 1290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  DB 202 TCATACTAAGGGCAGTGGCTAGACAGGAAAGAT 234

  RESULT 14
  A0478968/c
  LOCUS
  DEFINITION
    A0478968 562 bp DNA GSS 23-APR-1999
    RPCI-11-269K12.TV RPCI-11 Homo sapiens genomic clone RPCI-11-
    269K12, genomic survey sequence.
  ACCESSION
    A0478968
  VERSION
    A0478968.1 GI:4661087
  KEYWORDS
    GSS.
  SOURCE
    ORGANISM
      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 562)
    Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
    Venter,J.C.
    Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
    Map Building
    Unpublished (1997)
    On Dec 15, 1999 this sequence version replaced gi:4212149.
    Other GSS: RPCI-11-240H11.TV
    Contact: Shaying Zhao, William Nierman, Mark Adams
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: hbe@tigr.org
    Clones are derived from the human BAC library RPCI-11. For BAC
    library availability, please contact Pieter de Jong
    (pieter@dejong.med.buffalo.edu). Clones may be purchased from
    BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
    Research Genet cs (info@resgen.com). BAC end search page:
    http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
    Seq primer: SP6
    Class: BAC ends.
  FEATURES
    source
    location
    1..511
    /organism="Homo sapiens"
    /db_xref="GDB:7591954"
    /db_xref="taxon:9606"
    /clone="RPCI-11-240H11"
    /clone_lib="RPCI-11"
    /sex="Male"
    /cell_type="Lymphocytes"
    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
    RPCI11 Human Male BAC Library"
  BASE COUNT      181 a 79 c 89 g 162 t
  ORIGIN

```



Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
 cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbrrp/image/image.html](http://www-bio.llnl.gov/dbrrp/image/image.html)  
 Insert length: 908 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham.  
 Location/Qualifiers  
 1. 483  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_name="IMAGE:1372522"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="pH10P"

/note=Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-, provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TCGTACCAATCGATGGGAGCGCGCTCATTTTTTTTTTTT-3', (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was

```

Query Match          2.4%  Score 34.4;  DB 31;  Length 483;
Best Local Similarity 57.4%;  Pred. No. 6.7;
Matches 62;  Conservative 0;  Mismatches 46;  Indels 0;  Gaps 0;

QY  1199  ggtgaacattaccoggatgtagatcggggtaattgtaagtccaacaatcgctatttt 1258
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db    270  GATCATCTTGACCTCGAGTAGGCCCTAGGCTAATGTGTAAGTTTATGTCTTGTGTTTAAAC 211

QY  1259  aacagctaaagcaggtgcataatggggccagatacacccatcaatttg 1306
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db    210  AACACAAAAAAGGTATAAATGAAAAAAGAAAGAACTTTTAAAAATAG 163

```

Search completed: September 9, 2000, 20:45:53  
Job time: 2626 sec

---

(TM)

|    |      |      |     |   |        |                         |          |
|----|------|------|-----|---|--------|-------------------------|----------|
| 1  | 2393 | 99.0 | 328 | 1 | W40263 | K. oxytoca R-specific   | 1.62e+21 |
| 2  | 200  | 8.3  | 33  | 1 | W40375 | K. oxytoca R-specific   | 3.33e-07 |
| 3  | 157  | 6.5  | 21  | 1 | W40272 | K. oxytoca R-specific   | 1.05e-03 |
| 4  | 148  | 6.1  | 20  | 1 | W40264 | K. oxytoca R-specific   | 5.37e-03 |
| 5  | 132  | 5.5  | 18  | 1 | W40270 | K. oxytoca R-specific   | 9.25e-02 |
| 6  | 132  | 5.5  | 575 | 1 | W58856 | C. acidoovorans gamma-1 | 9.25e-02 |
| 7  | 122  | 5.0  | 19  | 1 | W40271 | K. oxytoca R-specific   | 5.23e-01 |
| 8  | 122  | 5.0  | 23  | 1 | W40273 | K. oxytoca R-specific   | 5.23e-01 |
| 9  | 107  | 4.4  | 482 | 1 | W08199 | B.licheniformis mutant  | 6.51e+00 |
| 10 | 107  | 4.4  | 561 | 1 | W97413 | Lucilla cuprina GABA r  | 6.51e+00 |
| 11 | 104  | 4.3  | 14  | 1 | W40274 | K. oxytoca R-specific   | 1.06e+01 |
| 12 | 102  | 4.2  | 483 | 1 | R22750 | Alpha amyrase.          | 1.47e+01 |
| 13 | 102  | 4.2  | 483 | 1 | R34289 | B.licheniformis alpha   | 1.47e+01 |
| 14 | 102  | 4.2  | 512 | 1 | W10325 | B.licheniformis acid    | 1.47e+01 |
| 15 | 100  | 4.1  | 483 | 1 | W08206 | B.licheniformis mutant  | 2.03e+01 |
| 16 | 100  | 4.1  | 483 | 1 | W05230 | Bacillus licheniformis  | 2.03e+01 |
| 17 | 100  | 4.1  | 483 | 1 | R58737 | M15E alpha-amyase.      | 2.03e+01 |
| 18 | 100  | 4.1  | 483 | 1 | R58721 | M197E alpha-amyase.     | 2.03e+01 |
| 19 | 100  | 4.1  | 483 | 1 | R58003 | M366A alpha-amyase.     | 2.03e+01 |
| 20 | 100  | 4.1  | 483 | 1 | W08193 | B.licheniformis mutant  | 2.03e+01 |
| 21 | 100  | 4.1  | 483 | 1 | R39742 | B.licheniformis alpha   | 2.03e+01 |
| 22 | 100  | 4.1  | 483 | 1 | W08211 | B.licheniformis mutant  | 2.03e+01 |
| 23 | 100  | 4.1  | 483 | 1 | W31404 | Bacillus licheniformis  | 2.03e+01 |

DR  
PTPT  
PTPT  
PTPT  
PTPT

WPI; 98-101063/09.  
Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation  
- by stereoselective hydrolysis of corresponding racemic amide using  
microorganism or derived enzyme, used as drug intermediate

### **IN CONFORMANCE WITH**

W40270;  
16-JUN-1998 (first entry)  
K. oxytoca R-specific amidohydrolase peptide T8.  
R-specific amidohydrolase; hydrolysis: racemate; nitrogen source;  
3,3',3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.  
Klebsiella oxytoca.  
W09801568-A2.  
15-JAN-1998.  
10-JUL-1997; E03670.  
03-MAR-1997; CH-000500.  
10-JUL-1996; CH-001723.  
(LONZ.) LONZA AG.  
Brieden W, Naughton A, Robins K, Shaw N, Tinschert A

Query Match 5.0%; Score 122; DB 1; Length 23;  
Best Local Similarity 91.3%; Pred. No. 5.23e-01;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



Query Match 4.2%; Score 102; DB 1; Length 483;  
Best Local Similarity 29.9%; Pred. No. 1.47e+01;

```
FT      /note= "Wild-type Met substituted by Thr"
FT      misc_difference 188
FT      /note= "Wild-type Asn substituted by Ser"
FT      misc_difference 209
FT      /note= "Wild-type Ala substituted by Val"
PN      W09639528-A2.
PD      12-DEC-1996.
PF      06-JUN-1996; U09089.
PR      06-JUN-1995; US-468700.
PA      (GEMV ) GENENCOR INT INC.
PI      Day A, Mitchinson C, Requaadt C, Ringer C, Ropp T;
PI      Solheim LP;
DR      WPI; 97-043150/04.
PT      Mutant alpha-amylase with improved low pH starch hydrolysis activity
PT      - has at least one asparagine residue substituted or deleted, useful
PT      in detergents
PS      Claim 7; Page -: 49pp; English.
CC      Novel mutant alpha-amylases contain a deletion or substitution of at
CC      least one asparagine residue from the wild-type. The present sequence
CC      is a specific example of such a mutant alpha-amylase in which the
CC      Met residue at position 15 has been substituted by a Thr residue, the
CC      Asn residue at position 188 has been substituted by a Ser residue, and
CC      the Ala residue at position 209 has been substituted by a Val residue.
CC      The mutant alpha-amylase is useful for starch processing and
CC      liquefaction. It is used in dish-washing or fabric-cleaning detergent
CC      compositions. The modifications produce enzymes with enhanced low pH
CC      performance suited for commercial liquefaction conditions. The mutant
CC      alpha-amylases are also found to have improved thermal and oxidative
CC      stability and to be more stable in the absence of calcium ions, compared
CC      to wild-type.
CC      N.B. the present sequence is not shown in the specification but is
CC      derived from the Bacillus licheniformis alpha-amylase sequence shown in
CC      figure 4.
SQ      Sequence 483 AA;

Query Match 4.1%; Score 100; DB 1; Length 483;
Best Local Similarity 29.9%; Pred. No. 2.03e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;

Db 69 QKGIVRTKYGTKELOSA-IKSLHSRDINVYGD-VVINHKGADATEDVTAVEVDP-ADR 125
QY 84 QNGPIWVNGAEKGDVLAVYIESMLPRGVDPYICAMIPHFGLTGT-DLTAMLDPLPEK 142
Db 126 NRVISGEHLIKAWTHFH 142
QY 143 VRMIKLDSEKVVWSKRH 159
```

Search completed: Thu Aug 24 18:54:59 2000  
Job time : 29 secs.









CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/704,706A  
CC FILING DATE: February 20, 1998  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Stone, Christopher  
CC REGISTRATION NUMBER: 35,696  
CC REFERENCE/DOCKET NUMBER: GC275-3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (650) 846-7555  
CC TELEFAX: (650) 845-6405  
CC INFORMATION FOR SEQ ID NO: 34:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 483 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 483 AA; 55211 MW; 1221290 CN;  
SQ  
Query Match 4.1%; Score 100; DB 2; Length 483;  
-Best Local Similarity 29.9%; Pred. No. 1.17e+01;  
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;  
Db 69 QKGVTRTKYGTGKELQSA-IKSLHSDINVGD-VVINHKGADATEDVTAVEVDP-ADR 125  
QY 84 QNGPIWNGAEKGDVLAVYIESMLPRGVDPYGCAMIPHFGGLTGT-DLTAMLNDPLPEK 142  
Db 126 NRVISGEHLIKAWTHF 142  
QY 143 VRMIKLDSEKVVWSKRH 159  
RESULT 7  
ID US-08-468-698-36 STANDARD; PRT; 483 AA.  
XX XXXXXX  
AC  
XX  
DT  
XX  
DE  
XX  
XX  
Sequence 36, Application US/08468698  
Sequence 36, Application US/08468698  
Patent No. 5849549  
GENERAL INFORMATION:  
CC APPLICANT: Antrim, Richard L.  
CC APPLICANT: Barnett, Christopher  
CC APPLICANT: Mitchinson, Colin  
CC APPLICANT: Power, Scott D.  
CC APPLICANT: Requaadt, Carol  
CC APPLICANT: Solheim, Leif P.  
CC TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase  
CC NUMBER OF SEQUENCES: 68  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Genencor International, Inc.  
CC STREET: 180 Kimball Way  
CC CITY: South San Francisco  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94080  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5 inch, 1.44 Mb  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/468,698  
CC FILING DATE: 06-JUN-95

CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/194,664  
CC FILING DATE: 10-FEB-94  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/016,395  
CC FILING DATE: 11-FEB-93  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Stone, Christopher L.  
CC REGISTRATION NUMBER: 35,696  
CC REFERENCE/DOCKET NUMBER: GC220D2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 742-7555  
CC TELEFAX: (415) 742-7217  
CC INFORMATION FOR SEQ ID NO: 36:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 483 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 483 AA; 55181 MW; 1222078 CN;  
SQ  
Query Match 4.1%; Score 100; DB 2; Length 483;  
-Best Local Similarity 29.9%; Pred. No. 1.17e+01;  
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;  
Db 69 QKGVTRTKYGTGKELQSA-IKSLHSDINVGD-VVINHKGADATEDVTAVEVDP-ADR 125  
QY 84 QNGPIWNGAEKGDVLAVYIESMLPRGVDPYGCAMIPHFGGLTGT-DLTAMLNDPLPEK 142  
Db 126 NRVISGEHLIKAWTHF 142  
QY 143 VRMIKLDSEKVVWSKRH 159  
RESULT 8  
ID US-08-890-383-3 STANDARD; PRT; 483 AA.  
XX XXXXXX  
AC  
XX  
DT  
XX  
DE  
XX  
XX  
Sequence 3, Application US/08890383  
Sequence 3, Application US/08890383  
Patent No. 6008026  
GENERAL INFORMATION:  
CC APPLICANT: Anthony G. Day  
CC TITLE OF INVENTION: MUTANT ALPHA-AMYLASE HAVING INTRODUCED  
CC TITLE OF INVENTION: THEREIN A DISULFIDE BOND  
CC NUMBER OF SEQUENCES: 11  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Genencor International, Inc.  
CC STREET: 925 Page Mill Road  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304-1013  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSeq for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/890,383  
CC FILING DATE: To Be Assigned  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Stone, Christopher  
CC REGISTRATION NUMBER: 35,696  
CC REFERENCE/DOCKET NUMBER: GC377  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (650) 846-7555

Query Match 4.1%; Score 100; DB 4; Length 483;  
Best Local Similarity 29.9%; Pred. No. 1.17e+01;  
Matches 23; Conservative 21; Mismatches 29; Indels

Query Match 4.1%; Score 100; DB 4; Length 483;  
Best Local Similarity 29.9%; Pred. No. 1.17e+01;  
Matches 23; Conservative 21; Mismatches 29; Indels

AC  
XXXXXXAC  
XXXXXX

```
DT XX Sequence 32, Application PC/TUS9401553A
DE XX Sequence 32, Application PC/TUS9401553A
XX XX GENERAL INFORMATION:
CC CC APPLICANT: GENENCOR INTERNATIONAL, INC.
CC CC TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
CC CC NUMBER OF SEQUENCES: 68
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Genencor International, Inc.
CC CC STREET: 180 Kimball Way
CC CC CITY: South San Francisco
CC CC STATE: CA
CC CC COUNTRY: USA
CC CC ZIP: 94080
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Floppy disk
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: PCT/US94/01553A
CC CC FILING DATE:
CC CC CLASSIFICATION:
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: Horn, Margaret A.
CC CC REGISTRATION NUMBER: 33,401
CC CC REFERENCE/DOCKET NUMBER: GC220-2
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: (415) 742-7536
CC CC TELEFAX: (415) 742-7217
CC CC INFORMATION FOR SEQ ID NO: 32:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 483 amino acids
CC CC TYPE: amino acid
CC CC STRANDEDNESS: single
CC CC TOPOLOGY: linear
CC CC MOLECULE TYPE: protein
CC CC SEQUENCE 483 AA; 55211 MW; 1221290 CN;

Query Match 4.1%; Score 100; DB 4; Length 483;
Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;

Db 69 QKGTVRTKYGKGELOSA-IKLSHRDINVYGD-VVINHGKGDATEDVTAVEVDP-ADR 125
QY 84 QNGPIWNGAEKGDVLAVIESMLPRGVDPYGICAMIPHFGGLTGT-DITAMNDPLPEK 142
Db 126 NRVISGEHLIKAWTHFH 142
QY 143 VRMIKLDSEKVVYWSKRH 159

RESULT 12
ID PCT-US95-10426-36 STANDARD; PRT; 483 AA.
XX AC xxxxxx
XX DT
XX DE
XX DE Sequence 36, Application PC/TUS9510426
XX XX Sequence 36, Application PC/TUS9510426
CC CC GENERAL INFORMATION:
CC CC APPLICANT: GENENCOR INTERNATIONAL, INC.
CC CC TITLE OF INVENTION: An Improved Cleaning Composition
CC CC NUMBER OF SEQUENCES: 68
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Genencor International
CC CC STREET: 180 Kimball Way
CC CC CITY: South San Francisco
CC CC STATE: CA

Sequence 32, Application PC/TUS9401553A
GENERAL INFORMATION:
APPLICANT: GENENCOR INTERNATIONAL, INC.
TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01553A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
REFERENCE/DOCKET NUMBER: GC220-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7536
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 483 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 483 AA; 55211 MW; 1221290 CN;

Query Match 4.1%; Score 100; DB 4; Length 483;
Best Local Similarity 29.9%; Pred. No. 1.17e+01;
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;

Db 69 QKGTVRTKYGKGELOSA-IKLSHRDINVYGD-VVINHGKGDATEDVTAVEVDP-ADR 125
QY 84 QNGPIWNGAEKGDVLAVIESMLPRGVDPYGICAMIPHFGGLTGT-DITAMNDPLPEK 142
Db 126 NRVISGEHLIKAWTHFH 142
QY 143 VRMIKLDSEKVVYWSKRH 159

RESULT 13
ID US-08-468-220-32 STANDARD; PRT; 483 AA.
XX AC xxxxxx
XX DT
XX DE
XX DE Sequence 32, Application US/08468220
XX XX Sequence 32, Application US/08468220
CC CC Patent No. 5824532
CC CC GENERAL INFORMATION:
CC CC APPLICANT: Antrim, Richard L.
CC CC APPLICANT: Barnett, Christopher
CC CC APPLICANT: Mitchinson, Colin
CC CC APPLICANT: Power, Scott D.
CC CC APPLICANT: Requaft, Carol
CC CC APPLICANT: Solheim, Leif P.
CC CC TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
CC CC NUMBER OF SEQUENCES: 68
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Genencor International, Inc.
CC CC STREET: 180 Kimball Way
CC CC CITY: South San Francisco
CC CC STATE: CA
CC CC COUNTRY: USA
CC CC ZIP: 94080
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: 3.5 inch, 1.44 Mb
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CC CURRENT APPLICATION DATA:
```



CC INFORMATION FOR SEQ ID NO: 37:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 487 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SQ SEQUENCE 487 AA; 55495 MW; 1241380 CN;

Query Match 4.1%; Score 100; DB 4; Length 487;  
Best Local Similarity 29.9%; Pred. No. 1.17e+01;  
Matches 23; Conservative 21; Mismatches 29; Indels 4; Gaps 4;  
Db 73 QKGTVRTKYGTGGELOSA-IKSLHSRDINVYGD-VVINHGKGDATEDVTAVEVDP-ADR 129  
QY 84 QNGPINVNGAEKGDVLAVYIESMLPRGVDPYGICAMIPHFGGLTGT-DLTAMINDPLPEK 142  
Db 130 NRVISGEHLIKAWTHF 146  
QY 143 VRMIKLDSEKVIWKRH 159

Search completed: Thu Aug 24 18:55:43 2000  
Job time : 27 secs.





```
RESULT 2
ENTRY   #type complete
TITLE   hypothetical protein TM0119 - Thermotoga maritima (strain
        MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE     11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
        G72414
ACCESSIONS
REFERENCE A72200
#authors
        Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
        R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
        W.C.; Ketchum, K.A.; McDonald, L.; Otterback, T.R.; Malek,
        J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
        M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
        Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
        Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal
#title
        Evidence for lateral gene transfer between Archaea and
        Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession G72414
#status preliminary
#molecule_type DNA
#residues 285 #label ARN
#cross-references GB:AE001698; GB:AE000512; NID:g4980609; PID:g4980611;
        TIGR:TM0119
#experimental_source strain MSB8
GENETICS
#gene TM0119
SUMMARY #length 285 #molecular-weight 31196 #checksum 8059
Query Match 11.0%; Score 266; DB 2; Length 285;
Best Local Similarity 30.8%; Pred. No. 3.15e-27;
Matches 52; Conservative 40; Mismatches 72; Indels 5; Gaps 5;
Db 117 IPIHMGVIGVAPGEVPTGTARRHGNGMDTKET-TEIVTVHLFPVQEGALLALGDVH 175
Qy 161 LPYKPHITGLSVPSIDINSITPDNHBGNDMDVPDIPGSGITYPL-VRAPGGRFLFIGDAH 219
Db 176 ATMGDGEVCVACEPAKVVVVDYSKE-EIKWPVVFENDAYIIVSLPDIIEEALKVETR 234
Qy 220 ACQGDGEICGTAVEFASITTIKVDLIKWKQLSWPRMENAIMSGSARPLEDATRIAYR 279
Db 235 ETV-WFIQRRKTIPTDAYMLASVDVIGISOLVNPAKTAKARIPIYIF 282
Qy 280 DLIYWLVEDFGFEOW-DAYMLLSQCGKVRLLGNMVDPKYTVGAMLNKLL 327
RESULT 3
ENTRY   #type complete
TITLE   probable acetamidase APE0528 - Aeropyrum pernix (strain K1)
ORGANISM #formal_name Aeropyrum pernix
DATE     20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
        C72750
ACCESSIONS
REFERENCE A72450
#authors
        Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
        Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
        S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
        Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
        Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.;
        Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
        Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal
#title
        Complete genome sequence of an aerobic hyper-thermophilic
        Crenarchaeon, Aeropyrum pernix K1.
#cross-references MUID:99310339
#accession C72750
#status preliminary
#molecule_type DNA
#residues 1-377 #label KAW
        #cross-references DBJ:AP000059; NID:g5103911; PIDN:BAA79495.1;
        PID:dl043281; PID:g5104179
        ##experimental_source strain K1
GENETICS
#gene APE0528
SUMMARY #length 377 #molecular-weight 41052 #checksum 5419
Query Match 8.2%; Score 197; DB 2; Length 377;
Best Local Similarity 34.2%; Pred. No. 4.84e-15;
Matches 40; Conservative 27; Mismatches 47; Indels 3; Gaps 3;
Db 146 LGVMGVARSKELLKEIKERDLKRGFVLPPPPGAVPPREPVASGLRTIPPRENGG 205
Qy 131 LTAMLNDPLPKVRMIKIDSEKVVYWSKRHTLPYKPHIGTSLVSP-EIDSINSLTPDNHGG 189
Db 206 NLDVHRHSPGSKIFYPVPEGA-LFSVGDHAHYAQDGVCGTAIMGATATLRFQVI 261
Qy 190 NMDVPDIPGSGITYPLVRAPGRLF-IGDAHACQGDGEICGTAVEFASITTIKVDLI 245
RESULT 4
ENTRY   #type complete
TITLE   hypothetical protein PH1041 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE     14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
        E71097
ACCESSIONS
REFERENCE A71000
#authors
        Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
        Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
        Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
        Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
        Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
        A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
        Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal
#title
        DNA Res. (1998) 5:55-76
        Complete sequence and gene organization of the genome of a
        hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
        OT3.
#cross-references MUID:98344137
#accession E71097
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-389 #label KAW
#cross-references GB:AP000004; NID:g3236131; PID:dl031082; PID:g3257456
#experimental_source strain OT3
#note this accession replaces an interim accession for a
        sequence replaced by GenBank
GENETICS
#gene PH1041
SUMMARY #length 389 #molecular-weight 42005 #checksum 7333
Query Match 6.7%; Score 163; DB 2; Length 389;
Best Local Similarity 38.7%; Pred. No. 1.91e-09;
Matches 24; Conservative 17; Mismatches 21; Indels 0; Gaps 0;
Db 206 GHLDVDVSRGCAIVIAPKVVEGGVYAGDAHAMGDEAGHTTDTVAETVTVEVIKGL 265
Qy 189 GNMDVPDIPGSGITYPLVRAPGRLFLIGDAHACQGDGEICGTAVEFASITTIKVDLKNW 248
Db 266 NL 267
Qy 249 QL 250
RESULT 5
ENTRY   #type complete
TITLE   formamidase (EC 3.5.1.49) A - Methylophilus methylotrophus
ORGANISM #formal_name Methylophilus methylotrophus
DATE     04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change
        S74213
ACCESSIONS
REFERENCE S74213; S78115
```

```

REFERENCE S74213
#authors Wyborn, N.R.; Mills, J.; Williams, S.G.; Jones, C.W.
#journal Eur. J. Biochem. (1996) 240:314-322
#title Molecular characterisation of formamidase from Methylophilus
        methylophilus.
#cross-references MUID:96439058
#accession S74213
#molecule_type DNA
##residues 1-407 ##label WYB
##cross-references EMBL:X99632; NID:g1480104; PID:e256826; PID:g1480105
##experimental_source strain NCIB 10515
#accession S78115
#molecule_type protein
##residues 1-5;44-47;172-176;211-215 ##label WYC
GENETICS
#gene fmdA
#keywords hydrolase
SUMMARY
#length 407 #molecular-weight 44466 #checksum 2787
Query Match 6.5%; Score 157; DB 2; Length 407;
Best Local Similarity 45.3%; Pred. No. 1.69e-08;
Matches 34; Conservative 14; Mismatches 20; Indels 7; Gaps 6;
Db 210 EGARTVPPRHGGNCDIKNLTKSRVFPVYVKD-GG-LSMGDLHFSQGDGEITFCG-AI 266
QY 177 DSINSLTPDNHGGNMDVPDIPGSGITY-PL-VRAPGGRGLFTGDAHACQGDGEI--CGTAV 232
Db 267 EMAGYLDIKVGLIKD 281
QY 233 EFASITTIKVLDLIKN 247
RESULT 6
ENTRY A47696 #type complete
TITLE acetamidase - Mycobacterium smegmatis
ORGANISM #formal_name Mycobacterium smegmatis
DATE 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS A47696
REFERENCE A47696
#authors Mahenthiralingam, E.; Draper, P.; Davis, E.O.; Colston, M.J.
#journal J. Gen. Microbiol. (1993) 139:575-583
#title Cloning and sequencing of the gene which encodes the highly
        inducible acetamidase of Mycobacterium smegmatis.
#cross-references MUID:93232774
#contents NCTC 8159
#accession A47696
#status preliminary
#molecule_type DNA; protein
##residues 1-406 ##label MAH
##cross-references GB:X57175; GB:S59037; NID:g312080; PID:g312081
#note sequence inconsistent with the nucleotide translation
        sequence extracted from NCBI backbone (NCBIN:129965,
        NCBIP:129966)
SUMMARY #length 406 #molecular-weight 43964 #checksum 926
Query Match 6.1%; Score 147; DB 2; Length 406;
Best Local Similarity 43.8%; Pred. No. 5.95e-07;
Matches 32; Conservative 14; Mismatches 22; Indels 5; Gaps 4;
Db 218 DGARTVPPRHGGNHDIKNFTGRGRIFFP-VFVEGAMSLGSDLHFSQGDGEINFCG-AIE 275
QY 177 DSINSLTPDNHGGNMDVPDIPGSGS-ITYPLVRAPGGRGLFTGDAHACQGDGEI--CGTAVE 233
Db 276 MGGFIDMHVDLIK 288
QY 234 FASITTIKVLDLIK 246
RESULT 7
ENTRY T04712 #type complete
TITLE probable formamidase (EC 3.5.1.49) F19F18.40 - Arabidopsis
        thaliana

```

```

ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
        cress
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
        11-Jun-1999
ACCESSIONS T04712
REFERENCE Z15382
#authors Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.;
        Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, March 1999
#accession T04712
#molecule_type DNA
##residues 1-432 ##label BEV
##cross-references EMBL:AL035605
##experimental_source cultivar Columbia; BAC clone F19F18
GENETICS
#map_position 4
#introns 73/3; 153/1; 196/3; 275/2; 356/3; 392/3
#note F19F18.40
#keywords hydrolase
SUMMARY
#length 432 #molecular-weight 47566 #checksum 5561
Query Match 5.6%; Score 135; DB 2; Length 432;
Best Local Similarity 35.6%; Pred. No. 3.72e-05;
Matches 26; Conservative 19; Mismatches 25; Indels 3; Gaps 2;
Db 210 EAARTIPGRENGGNODIKNLRSKIYLPVFVEGANLSTGDMHFSQGDGEISFCG-AIEM 268
QY 177 DSINSLTPDNHGGNMDVPDIPGSGITYPLVRAPGGRGLFTGDAHACQGDGEI--CGTAVEF 234
Db 269 SGFLEKCEIIRN 281
QY 235 ASITTIKVLDLIKN 247
RESULT 8
ENTRY B70434 #type complete
TITLE polysialic acid capsule expression protein - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
        29-Sep-1999
ACCESSIONS B70434
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
        Lennox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
        Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
        J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
        Aquifex aeolicus.
#cross-references MUID:98196666
#accession B70434
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
##residues 1-322 ##label AQF
##cross-references GB:AE000745; NID:g2983907; PIDN:AAC07460.1;
        PID:g2983910; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene kpsF
#CLASSIFICATION #superfamily probable ATP-binding protein gutQ; CBS homology
FEATURE
208-256 #domain CBS homology #label CBS1\
273-321 #domain CBS homology #label CBS2
SUMMARY #length 322 #molecular-weight 35114 #checksum 6980
Query Match 4.3%; Score 103; DB 1; Length 322;
Best Local Similarity 25.2%; Pred. No. 8.78e-01;
Matches 28; Conservative 35; Mismatches 41; Indels 7; Gaps 7;
Db 83 GDMGLLSDKDALI-AISNSGSGSTEVLVLYQYAKALNIPVIGITGNEKSS-LAKYSDVVLK 140
QY 50 GDRITVDTRDAFEAGINSEQDIPSQLKMPFLNPQNGPIM-VNGAEKGVLAIVTESMLP 108

```

42

GENETICS

#gene  
#map\_position 37 min  
CLASSIFICATION  
KEYWORDS  
FEATURE  
1-35  
36-326

#domain signal sequence #status predicted #label SIG\  
#product vitamin B12 transport permease protein btuC  
#status predicted #label MAT\  
#domain transmembrane #status predicted #label TM01\  
#domain transmembrane #status predicted #label TM02\  
#domain transmembrane #status predicted #label TM03\  
#domain transmembrane #status predicted #label TM04\  
#domain transmembrane #status predicted #label TM05\  
#domain transmembrane #status predicted #label TM06\  
#domain transmembrane #status predicted #label TM07\  
#domain transmembrane #status predicted #label TM08  
#length 326 #molecular-weight 34949 #checksum 1332

SUMMARY

Query Match 4.2%; Score 101; DB 1; Length 326;  
Best Local Similarity 30.0%; Pred. No. 1.55e+00;  
Matches 9; Conservative 12; Mismatches 8; Indels 1; Gaps 1;

Db 167 STSVDLRLQWYMMGGFGVDWQSWMLA 196  
QY 273 ATRIAYRDLIWLVEDFGQW-DAYMLLS 301

RESULT 12

ENTRY H69145 #type complete  
TITLE sensory transduction histidine kinase - Methanobacterium  
ORGANISM thermoautotrophicum (strain Delta H)  
#formal\_name Methanobacterium thermoautotrophicum  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change  
05-Jun-1998

ACCESSIONS H69145  
REFERENCE A69000

. #authors

#journal J. Bacteriol. (1997) 179:7135-7155  
#title Complete genome sequence of Methanobacterium  
thermoautotrophicum Delta H: functional analysis and  
comparative genomics.  
#cross-references MUID:98037514  
#accession H69145

##status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-567 ##label MTH  
##cross-references GB:A6000821; GB:A6000666; NID:g2621414; PID:g2621415  
##experimental\_source strain Delta H

GENETICS

#gene MTH356  
#start\_codon GTG  
SUMMARY #length 567 #molecular-weight 62431 #checksum 5446

Query Match 4.2%; Score 102; DB 2; Length 567;  
Best Local Similarity 25.2%; Pred. No. 1.17e+00;  
Matches 40; Conservative 49; Mismatches 55; Indels 15; Gaps 12;

Db 397 IQAMNASEAREVLRDAQSVRAMAILH-ET--IYDSGNFTGYDMSGFTRLRLERSVAY 453  
QY 55 VDTDAFEAGINEQDIPSLQKMPFNQNGIMPVNGAEKGDVLAVYTESMLPRGVDPI 114  
Db 454 GVYG-I-HFR-V-DADVRNLETAIPLGL-LIN-EA--VTNSIRHAPP-SGE-GSITVTM 503

QY 115 GICAMIPHFGGLTGDTLAMLNDPLPEKVRMIKLDSEKVVSKRHTLPYKPHIGTSLVSP 174

Db 504 ESDGLLYLRVEDDGTGME--GIPDGTGLSLMALADOL 540  
QY 175 EIDSINSLTPDNHGGMDVDPDIPGSGITYPLVRAPGRL 213

RESULT 13

ENTRY A64742 #type complete  
TITLE hypothetical protein b0177 - Escherichia coli  
ORGANISM #formal\_name Escherichia coli  
DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change  
29-Sep-1999

ACCESSIONS A64742  
REFERENCE A64720

#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;  
Kirpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
Y.

#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references MUID:97426617  
#accession A64742

##status nucleic acid sequence not shown; translation not shown  
##molecule\_type DNA  
##residues 1-810 ##label BLAT  
##cross-references GB:A6000127; GB:U00096; NID:g1786370;  
PIDN:AAC73288.1; PID:g1786374; UMG:P.b0177

##experimental\_source strain K-12, substrain MG1655  
CLASSIFICATION #superfamily protective surface antigen D-15  
SUMMARY #length 810 #molecular-weight 90552 #checksum 5454

Query Match 4.2%; Score 101; DB 2; Length 810;  
Best Local Similarity 27.1%; Pred. No. 1.55e+00;  
Matches 19; Conservative 22; Mismatches 24; Indels 5; Gaps 5;

Db 114 LEASGV-RVGESLDRITTIAD-IEKGLD-FDYISGVKYSASVAVVPLPRNRVDLKLVFQ 170 -  
QY 4 LEESIMAKRGVAGRKPVTHLTTEMQKEFHYTIGPSTPLVTI-EPGDRIIVDTDAFE 62

Db 171 EGVSAEIQOI 180  
QY 63 GAINSE-QDI 71

RESULT 14

ENTRY A44945 #type complete  
TITLE 104K microneme-rhoptry protein - Theileria parva  
ORGANISM #formal\_name Theileria parva  
DATE 14-May-1993 #sequence\_revision 14-May-1993 #text\_change  
09-Sep-1997

ACCESSIONS A44945  
REFERENCE A44945

#authors Iams, K.P.; Young, J.R.; Nene, V.; Desai, J.; Webster, P.;  
ole-Moiyoi, O.K.; Musoke, A.J.  
#journal Mol. Biochem. Parasitol. (1990) 39:47-60  
#title Characterisation of the gene encoding a 104-kilodalton  
microneme-rhoptry protein of Theileria parva.  
#cross-references MUID:90158697  
#accession A44945

##status preliminary  
##molecule\_type DNA  
##residues 1-924 ##label IAM  
##cross-references GB:M29954; NID:gl61865; PID:gl61866  
#length 924 #molecular-weight 103625 #checksum 2611

SUMMARY

Query Match 4.2%; Score 102; DB 2; Length 924;  
Best Local Similarity 22.3%; Pred. No. 1.17e+00;  
Matches 22; Conservative 30; Mismatches 45; Indels 5; Gaps 5;

Db 428 GLDNFLVCAMIVYVSDGVASL-VHLRIKRIIPANNDIYVLKGD-LYWTIRITKIOFTQEIKR 485

```

QY 110 GVDPYGICAMPHFGGLTGTDLTAMLNDPLPEKVRMIKLDSEKVVYWSKRHTLPYKPHIGT 169
Db 486 L-VKSKKKLAPITEEDSKHDEPPE-GPGASGLP-PKAPGDK 525
QY 170 LSVSPEDSINSUTPDNHGNDVDPDIGSGITYPLVRAPGGR 212

RESULT 15
ENTRY A37113 #type complete
TITLE ryanodine receptor, cardiac muscle - rabbit
ALTERNATE_NAMES ryanodine receptor_2
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 15-Feb-1991 #sequence_revision 15-Feb-1991 #text_change
20-Aug-1999
ACCESSIONS A37113; S74212
REFERENCE A37113
#authors Otsu, K.; Willard, H.F.; Khanna, V.K.; Zorzato, F.; Green,
N.M.; MacLennan, D.H.
#journal J. Biol. Chem. (1990) 265:13472-13483
#title Molecular cloning of cDNA encoding the Ca(2+) release channel
(ryanodine receptor) of rabbit cardiac muscle sarcoplasmic
reticulum
#cross-references MUID:90337947
#accession A37113
##status preliminary
##molecule_type mRNA
##residues 1-4969 ##label OTS
##cross-references GB:M59743; GB:J05564; NID:g164831; PIDN:AAA31179.1;
PID:g164832
REFERENCE S74212
#authors Nishida, K.; Otsu, K.; Hori, M.; Kuzuya, T.; Tada, M.
#journal Eur. J. Biochem. (1996) 240:408-415
#title Cloning and characterization of the 5'-upstream regulatory
region of the Ca2+-release channel gene of cardiac
sarcoplasmic reticulum.
#cross-references MUID:96439071
#accession S74212
##molecule_type DNA
##residues 1-16 ##label NIS
##cross-references EMBL:X99486; NID:e1000328
##experimental_source strain New Zealand White
GENETICS RYR2
#gene
CLASSIFICATION #superfamily ryanodine receptor; transcription initiation
factor sigma region 1 homology
KEYWORDS cardiac muscle; heart; phosphoprotein; transmembrane protein
SUMMARY #length 4969 #molecular-weight 565069 #checksum 5421
Query Match 4.2%; Score 101; DB 2; Length 4969;
Best Local Similarity 26.6%; Pred. No. 1.55e+00;
Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 3;
Db 3386 LKEPTPEAEELFRMVAEYFYWSKSHN--FKREEQNFVQNEINNMSPLITDTKS-KMSK 3442
QY 135 LNDPLPEKVRMIKLDSEK-VYWSKRHTLPYKPHIGTILSVSPEDSINSUTPDNHGNDV 193
Db 3443 AAVS 3446
QY 194 PDIG 197

```

Search completed: Thu Aug 24 18:54:13 2000  
Job time : 54 secs.

\*\*\*\*\*

W O R L D

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Aug 24 18:51:08 2000; MasPar time 11.48 Seconds  
Tabular output not generated. 885.716 Million cell updates/sec

Title: >US-09-214-679-2

Description: (1-328) from US09214679.pap

Sequence: 1 MKWLESIIMAKRGVAGRKP.....GNMVDPKYTVGAMLNKLLV 328

Scoring table: PAM 150  
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 49.160; Variance 85.522; scale 0.575

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                         | Pred. No. |
|------------|-------|-------------|--------|----|-------------------------------------|-----------|
| 1          | 157   | 6.5         | 407    | 1  | FMDA_METME FORMAMIDASE (EC 3.5.1.1) | 6.09e-10  |
| 2          | 147   | 6.1         | 405    | 1  | AMDA_MYCSM ACETAMIDASE (EC 3.5.1.1) | 3.27e-08  |
| 3          | 125   | 5.2         | 375    | 1  | FTS2_HALSA CELL DIVISION PROTEIN    | 1.33e-04  |
| 4          | 110   | 4.6         | 498    | 1  | MEFA_MOUSE MYOCTE-SPECIFIC ENHANCER | 2.45e-02  |
| 5          | 103   | 4.3         | 322    | 1  | YF46_AQUAE HYPOTHETICAL PROTEIN A   | 2.40e-01  |
| 6          | 101   | 4.2         | 326    | 1  | BTUC_ECOLI VITAMIN B12 TRANSPORT    | 4.50e-01  |
| 7          | 102   | 4.2         | 512    | 1  | AMY_BACLI ALPHA-AMYLASE PRECURSOR   | 3.29e-01  |
| 8          | 101   | 4.2         | 810    | 1  | UP05_ECOLI UNKNOWN PROTEIN FROM 2   | 4.50e-01  |
| 9          | 102   | 4.2         | 924    | 1  | 104K_THEPA 104 KDA MICRONEME-RHOP   | 3.29e-01  |
| 10         | 101   | 4.2         | 4969   | 1  | RYNC_RABIT RYANODINE RECEPTOR, CA   | 4.50e-01  |
| 11         | 99    | 4.1         | 181    | 1  | ARF_CRYNE ADP-RIBOSYLATION FACTO    | 8.38e-01  |
| 12         | 98    | 4.1         | 380    | 1  | FTS2_MYCPN CELL DIVISION PROTEIN    | 1.14e+00  |
| 13         | 98    | 4.1         | 396    | 1  | CHSB_IPOPU CHALCONE SYNTHASE B (E   | 1.14e+00  |
| 14         | 99    | 4.1         | 397    | 1  | CAPB_BACAN CAPB PROTEIN             | 8.38e-01  |
| 15         | 98    | 4.1         | 460    | 1  | FD3C_RICCO OMEGA-3 FATTY ACID DES   | 1.14e+00  |
| 16         | 98    | 4.1         | 632    | 1  | RCAC_FREDI RCAC PROTEIN             | 1.14e+00  |
| 17         | 99    | 4.1         | 5032   | 1  | RYNR_HUMAN RYANODINE RECEPTOR, SK   | 8.38e-01  |
| 18         | 99    | 4.1         | 5035   | 1  | RYNR_PIG RYANODINE RECEPTOR, SK     | 8.38e-01  |
| 19         | 96    | 4.0         | 135    | 1  | PILG_PSEAE PILG PROTEIN             | 2.08e+00  |
| 20         | 97    | 4.0         | 140    | 1  | JANE_DROME SEX-REGULATED PROTEIN    | 1.54e+00  |
| 21         | 96    | 4.0         | 180    | 1  | ARF1_ARATH ADP-RIBOSYLATION FACTO   | 2.08e+00  |
| 22         | 96    | 4.0         | 180    | 1  | ARF_ORYSA ADP-RIBOSYLATION FACTO    | 2.08e+00  |
| 23         | 96    | 4.0         | 180    | 1  | ARF_MAIZE ADP-RIBOSYLATION FACTO    | 2.08e+00  |

|    |    |     |      |   |                                   |          |
|----|----|-----|------|---|-----------------------------------|----------|
| 24 | 96 | 4.0 | 180  | 1 | ARF1_CHLRE ADP-RIBOSYLATION FACTO | 2.08e+00 |
| 25 | 97 | 4.0 | 500  | 1 | ABFA_BACSU TRIPEPTIDYL-PEPTIDASE  | 1.54e+00 |
| 26 | 96 | 4.0 | 1249 | 1 | TPP2_HUMAN TRIPETIDYL-PEPTIDASE   | 2.08e+00 |
| 27 | 97 | 4.0 | 1853 | 1 | CIPA_CLOTM CELLULOSONAL SCAFFOLDI | 1.54e+00 |
| 28 | 95 | 3.9 | 340  | 1 | CARP_POLTU POLYUROSOPESIN (EC 3.4 | 2.81e+00 |
| 29 | 95 | 3.9 | 396  | 1 | CHSA_IPOPU CHALCONE SYNTHASE A (E | 2.81e+00 |
| 30 | 95 | 3.9 | 417  | 1 | IE63_HSVSA 52 KDA IMMEDIATE-EARLY | 2.81e+00 |
| 31 | 95 | 3.9 | 446  | 1 | FD3C_ARATH OMEGA-3 FATTY ACID DES | 2.81e+00 |
| 32 | 94 | 3.9 | 498  | 1 | YHJ1_ECOLI PROTEIN YHJ1 PRECURSOR | 3.78e+00 |
| 33 | 95 | 3.9 | 605  | 1 | VE1_HPV09 REPLICATION PROTEIN E1  | 2.81e+00 |
| 34 | 95 | 3.9 | 1262 | 1 | TPP2_MOUSE TRIPEPTIDYL-PEPTIDASE  | 2.81e+00 |
| 35 | 95 | 3.9 | 3390 | 1 | POLG_DEN3 GENOME POLYPROTEIN (CO  | 2.81e+00 |
| 36 | 92 | 3.8 | 180  | 1 | ARF1_DAUCA ADP-RIBOSYLATION FACTO | 6.77e+00 |
| 37 | 93 | 3.8 | 182  | 1 | ARF_AJECA ADP-RIBOSYLATION FACTO  | 5.06e+00 |
| 38 | 92 | 3.8 | 182  | 1 | FTN1_HAEIN FERRITIN LIKE PROTEIN  | 6.77e+00 |
| 39 | 93 | 3.8 | 363  | 1 | CHSB_IPOCO CHALCONE SYNTHASE B (E | 5.06e+00 |
| 40 | 92 | 3.8 | 373  | 1 | YNS2_YEAST HYPOTHETICAL 41.2 KDA  | 6.77e+00 |
| 41 | 93 | 3.8 | 560  | 1 | MIS_HUMAN MUELLERIAN INHIBITING   | 5.06e+00 |
| 42 | 93 | 3.8 | 772  | 1 | CIPB_CLOTM CELLULOSONAL SCAFFOLDI | 5.06e+00 |
| 43 | 93 | 3.8 | 1010 | 1 | CONT_CHICK CONTACTIN PRECURSOR (N | 5.06e+00 |
| 44 | 92 | 3.8 | 1068 | 1 | SPS_MAIZE SUCROSE-PHOSPHATE SYNT  | 6.77e+00 |
| 45 | 92 | 3.8 | 1105 | 1 | YEGE_ECOLI HYPOTHETICAL 123.9 KDA | 6.77e+00 |

ALIGNMENTS

| RESULT | ID   | FMDA_METME                        | STANDARD; | PRT; | 407 AA. |
|--------|--|-----------------------------------|-----------|------|---------|
| AC     | Q50228;  |                                   |           |      |         |
| DT     | 15-DEC-1998  | (Rel. 37, Created)                |           |      |         |
| DT     | 15-DEC-1998  | (Rel. 37, Last sequence update)   |           |      |         |
| DT     | 15-DEC-1998  | (Rel. 37, Last annotation update) |           |      |         |
| DE     | FORMAMIDASE (EC 3.5.1.49)  | (FORMAMIDE AMIDOHYDROLASE).       |           |      |         |
| GN     | FMDA.  |                                   |           |      |         |
| OS     | Methylophilus methylotrophus (Bacterium W3A1).                             |                                   |           |      |         |
| OC     | Bacteria; Proteobacteria; beta subdivision; Methylophilus group;           |                                   |           |      |         |
| OC     | Methylophilus.   |                                   |           |      |         |
| RN     | [1]  |                                   |           |      |         |
| RP     | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.                                  |                                   |           |      |         |
| RC     | STRAIN-NCIB 10515;   |                                   |           |      |         |
| RX     | MEDLINE: 96439058;   |                                   |           |      |         |
| RA     | Wyborn N.R., Mills J., Williams S.G., Jones C.W.;                          |                                   |           |      |         |
| RT     | "Molecular characterisation of formamidase from Methylophilus              |                                   |           |      |         |
| RT     | methylotrophus.";  |                                   |           |      |         |
| RL     | Eur. J. Biochem. 240:314-322(1996).  |                                   |           |      |         |
| RN     | [2]  |                                   |           |      |         |
| RP     | CHARACTERIZATION, AND SEQUENCE OF 1-5.                                     |                                   |           |      |         |
| RC     | STRAIN-NCIB 10515;   |                                   |           |      |         |
| RA     | Wyborn N.R., Scherr D.J., Jones C.W.;                                      |                                   |           |      |         |
| RT     | "Purification, properties and heterologous expression of formamidase       |                                   |           |      |         |
| RL     | from Methylophilus methylotrophus.";                                       |                                   |           |      |         |
| CC     | Microbiology 140:191-195(1994).  |                                   |           |      |         |
| CC     | -!- FUNCTION: HYDROLYSE FORMAMIDE WITH THE PRODUCTION OF AMMONIA WHICH     |                                   |           |      |         |
| CC     | CAN BE USED AS A SOURCE OF NITROGEN FOR GROWTH. ALSO ACTS, MORE            |                                   |           |      |         |
| CC     | SLOWLY, ON ACETAMIDE, PROPANAMIDE AND BUTANAMIDE.                          |                                   |           |      |         |
| CC     | -!- CATALYTIC ACTIVITY: FORMAMIDE + H(2)O = FORMATE + NH(3).               |                                   |           |      |         |
| CC     | -!- SUBUNIT: HOMOTRIMER.   |                                   |           |      |         |
| CC     | -!- MASS SPECTROMETRY: MW=44481; MW_ERR=26; METHOD=ELECTROSPRAY.           |                                   |           |      |         |
| CC     | -!- SIMILARITY: STRONG, TO M.SMEGMATIS ACETAMIDASE.                        |                                   |           |      |         |
| CC     | -----  |                                   |           |      |         |
| CC     | This SWISS-PROT entry is copyright. It is produced through a collaboration |                                   |           |      |         |
| CC     | between the Swiss Institute of Bioinformatics and the EMBL outstation -    |                                   |           |      |         |
| CC     | the European Bioinformatics Institute. There are no restrictions on its    |                                   |           |      |         |
| CC     | use by non-profit institutions as long as its content is in no way         |                                   |           |      |         |
| CC     | modified and this statement is not removed. Usage by and for commercial    |                                   |           |      |         |
| CC     | entities requires a license agreement (See http://www.isb-sib.ch/announce/ |                                   |           |      |         |
| CC     | or send an email to license@isb-sib.ch).                                   |                                   |           |      |         |
| CC     | -----  |                                   |           |      |         |
| DR     | EMBL: X99632; CAA67953.1;  |                                   |           |      |         |
| KW     | Hydrolase.   |                                   |           |      |         |
| SQ     | SEQUENCE. 407 AA; 44466 MW; C55E888BA4C44A70 CRC64;                        |                                   |           |      |         |
|        | Query Match 6.5%; Score 157; DB 1; Length 407;                             |                                   |           |      |         |

```

Best Local Similarity 45.3%; Pred. No. 6.09e-10;
Matches 34; Conservative 14; Mismatches 20; Indels 7; Gaps 6;

Db 210 EGARTVPPRHGGNGCDIKNLTKGSRVPPVVKD-GG-LSMGDLHFSOGDGEITFCG-AI 266
QY 177 DSINSLTPDNHGGNMDVPDIPGGSITY-PL-VRAPEGRLFIGDAHACQDGEI--CGTAV 232
Db 267 EMAGYLDIKVGLIKD 281
QY 233 EFASITTIKVLDLIK 247

RESULT 2
ID AMDA_MYCSM STANDARD; PRT; 405 AA.
AC Q07838;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ACETAMIDASE (EC 3.5.1.-).
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23; 81-100 AND 180-194.
RC STRAIN-NCTC 8159;
RX MEDLINE; 93232774.
RA Mahenthiralingam E., Draper P., Davis E.O., Colston M.J.;
RT "Cloning and sequencing of the gene which encodes the highly
RT inducible acetamidase of Mycobacterium smegmatis.";
RL J. Gen. Microbiol. 139:575-583(1993).
CC -1- FUNCTION: ALLOWS ACETAMIDE TO BE USED AS A SOLE CARBON OR NITROGEN
CC SOURCE.
CC -1- CATALYTIC ACTIVITY: FORMAMIDE + H(2)O = FORMATE + NH(3).
CC -1- SIMILARITY: STRONG. TO M.METHYLOTROPHUS FORMAMIDASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X57175; CAA0462.1; -.
CC KW Hydrolase.
CC FT INIT_MET 0 0 POTENTIAL.
CC SQ SEQUENCE 405 AA; 43833 MW; 05E2E359210E0BF2 CRC64;

Query Match 6.1%; Score 147; DB 1; Length 405;
Best Local Similarity 43.8%; Pred. No. 3.27e-08;
Matches 32; Conservative 14; Mismatches 22; Indels 5; Gaps 4;

Db 217 DGARTVPPRENGNHDIKNETRGSRIYFP-VFVEGAMLSGGDLHFSOGDGEITFCG-AIE 274
QY 177 DSINSLTPDNHGGNMDVPDIPGGSITY-ITPLVRAPGRLFIGDAHACQDGEI--CGTAVE 233
Db 275 MGFIDMHVDLIK 287
QY 234 FASITTIKVLDLIK 246

RESULT 3
ID FTSZ_HALSA STANDARD; PRT; 375 AA.
AC Q48290;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CELL DIVISION PROTEIN FTSZ_HOMOLOG.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
RN [1]

```

```

SEQUENCE FROM N.A.
RP STRAIN-PH081;
RX MEDLINE; 96200101.
RA Maggolin W., Wang R., Kumar M.;
RT "Isolation of an ftsz homolog from the archaeobacterium Halobacterium
RT salinarum: implications for the evolution of FtsZ and tubulin.";
RL J. Bacteriol. 178:1320-1327(1996).
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.
CC ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF
CC THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,
CC AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPARATION TO
CC BEGIN. BINDS TO AND HYDROLYZES GTP (BY SIMILARITY).
CC -1- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32860; AAB06191.1; -.
CC DR HSSP; Q57816; IFSZ.
CC PFAM; PF00091; tubulin; 1.
CC PRINTS; PR00423; CELLDIVISFTSZ.
CC PROSITE; PS01134; FTSZ_1; 1.
CC PROSITE; PS01135; FTSZ_2; 1.
CC Cell division; Septation; GTP-binding.
CC FT NP_BIND 105 113 GTP (POTENTIAL).
CC SQ SEQUENCE 375 AA; 39261 MW; 43A05013AB3E5BEB CRC64;

Query Match 5.2%; Score 125; DB 1; Length 375;
Best Local Similarity 23.8%; Pred. No. 1.33e-04;
Matches 34; Conservative 35; Mismatches 65; Indels 9; Gaps 8;

Db 186 DQIAETVKGISSETIPQSLINLDYADMTAIMNCGVAVMLVGETQDKNKTNEVVKDAMN 245
QY 51 DRIIVDFDAFEAGINSQDIPSOQLKMPFLNPQNG-PIM-VNGA-EKGDVLAVIESML 107
Db 246 HPLDLDVYRGASGLVHTIG--GPDLTUKEAGIADNI-TERLDASANVIGARTQESYK 302
QY 108 PRGVD-PY-GICAMTPHFGLTGTDLTAMLNDPLPEKVRMIKLD-SKVYVWSKRHTLPYK 164
Db 303 GKVRVMAIMTGVQSAQVLGPSTQ 325
QY 165 PHIGTLVSPEIDSINSLTPDNH 187

RESULT 4
ID MEFA_MOUSE STANDARD; PRT; 498 AA.
AC Q60929;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOCYTE-SPECIFIC ENHANCER FACTOR 2A.
GN MEF2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-CEREbellum;
RX MEDLINE; 97165895.
RA Lin X., Shah S., Bulleit R.F.;
RT "The expression of MEF2 genes is implicated in CNS neuronal
RT differentiation.";
RL Brain Res. Mol. Brain Res. 42:307-316(1996).
CC -1- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE
CC MEF2 ELEMENT PRESENT IN THE REGULATORY REGIONS OF MANY MUSCLE-
CC SPECIFIC GENES. ACTIVATES TRANSCRIPTION VIA THIS ELEMENT. MAY BE
CC INVOLVED IN MUSCLE-SPECIFIC AND/OR GROWTH FACTOR-RELATED

```



CC TRANSCRIPTION (BY SIMILARITY).  
CC -!- SUBUNIT: HETERODIMER.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION  
CC FACTORS. MEF2 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U30823; AAA74030.1; -  
CC HSSP: P11831; 1SR5.  
CC MGD: MGI:99532; MEF2A.  
CC DR MGF: PF00319; SRF-TF; 1.  
CC DR PRAM: PR00404; MADS DOMAIN.  
CC DR PROSITE: PS00350; MADS\_BOX\_1; 1.  
CC DR PROSITE: PS00666; MADS\_BOX\_2; 1.  
CC KW Transcription regulation; Nuclear protein; DNA-binding; Activator;  
CC Multigene family.  
CC FT DOMAIN 3 57 MADS.  
CC FT DNA\_BIND 58 86 MEF2-TYPE (POTENTIAL).  
CC FT DOMAIN 254 257 POLY-PRO.  
CC FT DOMAIN 288 293 POLY-GLU.  
CC FT DOMAIN 419 423 POLY-HIS.  
CC FT DOMAIN 448 455 POLY-SER.  
CC SQ SEQUENCE 498 AA; 53724 MW; 590678D1BD1B3723 CRC64;  
Query Match 4.6%; Score 110; DB 1; Length 498;  
Best Local Similarity 29.2%; Pred. No. 2.45e-02;  
Matches 19; Conservative 20; Mismatches 25; Indels 1; Gaps 1;  
DB 107 HSPLED-REIKLNEUSDFTFKRGPCFPQNFMSVTVPTSPNLSLDTNPGSSLSVSPS 165  
QY 136 NDPLEKVRMIKLDSEKVVYMKRHTLPYKPHICTLSVPEIDSINSITPDNHHGNDVDP 195  
DB 166 LAASS 170  
QY 196 IGPSS 200  
RESULT 5  
ID YF46\_AQUAE STANDARD; PRT; 322 AA.  
AC O67500;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL PROTEIN AQ\_1546.  
GN AQ\_1546.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE: 98196666.  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358(1998).  
CC -!- SIMILARITY: BELONGS TO THE SIS FAMILY. GUTQ/KPSF SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: AE000745; AAC07460.1; -  
CC PFAM: PF00571; CBS; 2.  
CC DR PFAM: PF01380; SIS; 1.  
CC KW Hypothetical protein; ATP-binding; Repeat; CBS domain.  
CC FT NP\_BIND 50 55 ATP (POTENTIAL).  
CC FT DOMAIN 203 256 CBS 1.  
CC FT DOMAIN 268 321 CBS 2.  
CC SQ SEQUENCE 322 AA; 35114 MW; 74BE86D9B8FB2A69 CRC64;  
Query Match 4.3%; Score 103; DB 1; Length 322;  
Best Local Similarity 25.2%; Pred. No. 2.40e-01;  
Matches 28; Conservative 35; Mismatches 41; Indels 7; Gaps 7;  
DB 83 GDMGLDSKDALI-AISNGSESTVLVYQAKALNPVIGITGNEKSS-LAKYSDVVLK 140  
QY 50 GDRIIVTRDAFEGAINSEQDIPSQLKMPFNQNGPIM-VNGAEKGDVLAVIESMLP 108  
DB 141 IPVDR-EACP-F-NLAPTSTVTALGDAITAMTLKLGFSQEDF-AKRH 187  
QY 109 RGVDPYICAMIPHFGGLTGDTLAMLNDPLPEKVRMIKLDSEKVVYWSKRH 159  
RESULT 6  
ID BTUC\_ECOLI STANDARD; PRT; 326 AA.  
AC P06609; P7197;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE VITAMIN B12 TRANSPORT SYSTEM PERMEASE PROTEIN BTUC.  
GN BTUC.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 86304184.  
RA Friedrich M.J., Deveaux L.C., Kadner R.J.;  
RT "Nucleotide sequence of the btuCD genes involved in vitamin B12  
RT transport in Escherichia coli and homology with components of  
RT periplasmic-binding-protein-dependent transport systems";  
RL J. Bacteriol. 167:928-934(1986).  
RN [2]  
RP REVISIONS.  
RX MEDLINE: 89364713.  
RA Rioux C.R., Kadner R.J.;  
RT "Vitamin B12 transport in Escherichia coli K12 does not require the  
RT btuE gene of the btuCED operon";  
RL Mol. Gen. Genet. 217:301-308(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12/MG1655;  
RX MEDLINE: 97426617.  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:12453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horluchi T.,  
RA Ikemoto K., Inada T., Isono S., Itoh T., Kanai K., Kasai H.,  
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,  
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,  
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
CC FOR VITAMIN B12. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE  
CC SUBSTRATE ACROSS THE MEMBRANE.  
CC -----







DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CELL DIVISION PROTEIN FTSZ.  
 GN FTSZ.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 CC Mycoplasmataceae; Mycoplasma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE; 97105885  
 RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Herrmann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RT pneumoniae.";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.  
 CC ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF  
 CC THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,  
 CC AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPTATION TO  
 CC BEGIN. BINDS TO AND HYDROLYZES GTP (BY SIMILARITY).  
 CC -1- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSEMBLE AT THE INNER SURFACE  
 CC OF THE CYTOPLASMIC MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE FTSZ FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF000051; 'AAB96167.1'; -  
 DR PFAM; PF00091; tubulin; 1.  
 DR PROSITE; PS01134; FTSZ\_1; 1.  
 DR PROSITE; PS01135; FTSZ\_2; 1.  
 KW Cell division; Septation; GTP-binding.  
 FT NP\_BIND 115 123 GTP (POTENTIAL).  
 SQ SEQUENCE 380 AA; 42797 MW; 6C3B531C63F4997E CRC64;  
 -Query Match 4.1%; Score 98; DB 1; Length 380;  
 Best Local Similarity 31.5%; Pred. No. 1.14e+00;  
 Matches 17; Conservative 17; Mismatches 19; Indels 1; Gaps 1;  
 Db 82 VGGDPKQAVLAHFFLEQPHKLSDFDFCLVAGFGKGTGATPVSFSLSK 135  
 QY 90 VNG-AEKGDVLAIVIESMLPRGVDPYGCAMIPHFGLGTDLTAMNDLPLEK 142  
 RESULT 13  
 ID CHSB\_IPOPU STANDARD; PRT; 396 AA.  
 AC P48398;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CHALCONE SYNTHASE B (EC 2.3.1.74) (NARINGENIN-CHALCONE SYNTHASE B)  
 DE (CHS-B).  
 GN CHSB.  
 OS Ipomoea purpurea (Common morning-glory).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;  
 OC Convolvulaceae; Ipomoea.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95241498.  
 RA Durbin M.L., Learn G.H., Huttley G.A., Clegg M.T.;  
 RT "Evolution of the chalcone synthase gene family in the genus  
 RT Ipomoea.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3338-3342(1995).  
 CC -1- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-

CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)  
 CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO  
 CC NARINGENIN.  
 CC -1- CATALYTIC ACTIVITY: 3 MALONYL-COA + 4-COUMAROYL-COA = 4 COA +  
 CC NARINGENIN-CHALCONE + 3 CO(2).  
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF  
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY  
 CC OF WHICH ARE BRIGHTLY COLORED.  
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U15947; AAC49030.1; -  
 DR PFAM; PF00195; Chal\_stil\_synth; 1.  
 DR PROSITE; PS00441; CHALCONE\_SYNTH; 1.  
 KW Flavonoid biosynthesis; Transferase; Acyltransferase;  
 KW Multigene family.  
 FT ACT\_SITE 170 170 BY SIMILARITY.  
 SQ SEQUENCE 396 AA; 43926 MW; 352C478C1AF892B6 CRC64;  
 Query Match 4.1%; Score 98; DB 1; Length 396;  
 Best Local Similarity 23.7%; Pred. No. 1.14e+00;  
 Matches 18; Conservative 23; Mismatches 33; Indels 2; Gaps 2;  
 Db 3 TTVTV-LTDTWSRRAKLEGDAKILAGTATPASWDQTTYPFYFRITNSQHLEHKEK 61  
 QY 238 TTKVDLIKKNQWLSWPRMENAEMISGSRPLEDATRIAYRDLIYWLVE-DFGFEQMD 296.  
 Db 62 FRICNKSIRKRLV 77  
 QY 297 YMLLSQCGKVRLLGNV 312  
 RESULT 14  
 ID CAPB\_BACAN STANDARD; PRT; 397 AA.  
 AC P19580;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE CAPB PROTEIN.  
 GN CAPB.  
 OS Bacillus anthracis.  
 OG Plasmid pTE702.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89123146.  
 RA Makino S.I., Uchida I., Terakado N., Sasakawa C., Yoshikawa M.;  
 RT "Molecular characterization and protein analysis of the cap region,  
 RT which is essential for encapsulation in Bacillus anthracis.";  
 RL J. Bacteriol. 171:722-730(1989).  
 CC -1- FUNCTION: ESSENTIAL FOR THE SYNTHESIS OF THE POLYGLUTAMATE  
 CC CAPSULE OF BACILLUS ANTHRACIS. MAY FORM A POLYGLUTAMYL SYNTHETASE  
 CC COMPLEX TOGETHER WITH CAPA AND CAPC PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M24150; AAA22286.1; -  
 DR FIR; A30091; A30091.



(™)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Title: >US-09-214-679-2

Description: (1-328) from US09214679.ppt

Perfect Score: 2417  
Sequence: 1 MKWLEESIMAKRGVGAGRKP.....GNMVDPKYTYGCAMLNKNLLV 328

Scoring table: PAM 150

Gap 11

Searched: 225878 seqs, 69334122 residues

post-processing: Minimum Match 0%

Listing first 45 summaries

Database: sptrembl12

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 47.366; Variance 79.598; scale 0.595

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Query Match | Score | Query |        |        | ID                      | Description | Pred. No. |
|------------|-------------|-------|-------|--------|--------|-------------------------|-------------|-----------|
|            |             |       | Match | Length | DB     |                         |             |           |
| 1          | 266         | 11.0  | 285   | 2      | Q9WXX3 | ACETAMIDASE, PUTATIVE.  | 1.52e-32    |           |
| 2          | 197         | 8.2   | 377   | 1      | Q9YEG1 | 377AA LONG HYPOTHETICA  | 3.16e-18    |           |
| 3          | 163         | 6.7   | 389   | 1      | O58755 | 389AA LONG HYPOTHETICA  | 1.19e-11    |           |
| 4          | 110         | 4.6   | 2091  | 3      | P78211 | FATTY ACID SYNTHASE, B  | 1.64e-02    |           |
| 5          | 109         | 4.5   | 789   | 5      | Q26216 | SIMILARITY TO CYTOCHRO  | 2.34e-02    |           |
| 6          | 103         | 4.3   | 322   | 2      | O67500 | POLYSTALIC ACID CAPSUL  | 1.85e-01    |           |
| 7          | 103         | 4.3   | 393   | 2      | O53272 | NIFS-LIKE PROTEIN.      | 1.85e-01    |           |
| 8          | 103         | 4.3   | 436   | 10     | Q40118 | DELTA-15 LINEOYL DESAT  | 1.85e-01    |           |
| 9          | 105         | 4.3   | 473   | 2      | O51454 | TRANSCRIPTIONAL ACTIVA  | 9.38e-02    |           |
| 10         | 104         | 4.3   | 597   | 5      | O45218 | ALXLY-DIHYDROXYACETONE  | 1.32e-01    |           |
| 11         | 104         | 4.3   | 1874  | 2      | Q3765  | IGA1 PROTEASE (EC 3.4.  | 1.32e-01    |           |
| 12         | 102         | 4.2   | 243   | 5      | O17308 | GABA RECEPTOR SUBUNIT   | 2.60e-01    |           |
| 13         | 102         | 4.2   | 567   | 1      | O26456 | SENSORY TRANSDUCTION H  | 2.60e-01    |           |
| 14         | 101         | 4.2   | 4968  | 6      | Q29621 | CARDIAC RYANODINE RECE  | 3.63e-01    |           |
| 15         | 99          | 4.1   | 281   | 2      | O30796 | MXAE.                   | 7.03e-01    |           |
| 16         | 98          | 4.1   | 387   | 10     | Q9ZPT7 | OMEGA-3 FATTY ACID DES  | 9.75e-01    |           |
| 17         | 99          | 4.1   | 396   | 10     | O04970 | CHALCONE SYNTHASE (EC   | 7.03e-01    |           |
| 18         | 98          | 4.1   | 899   | 3      | O59868 | CA+++-TRANSPORTING ATPA | 9.75e-01    |           |
| 19         | 99          | 4.1   | 1056  | 10     | O22855 | HYPOTHETICAL PROTEIN.   | 7.03e-01    |           |
| 20         | 100         | 4.1   | 1061  | 5      | O96998 | LH156.2 PROTEIN.        | 5.06e-01    |           |

```

QY 220 ACQGDGEICGTAVEFASITTIKVDLIKNNQLSWPRMENAINMSIGSARPLEDATEIAYR 279
Db 235 ETY-WFIQRKKTPTFTDAYMLASLSDVGVISQVLPNPAKTAKARIPKYIF 282
QY 280 DLIYWLVEDFGEQW-DAYMILLSQCGKVRGLGNMVDPKYTVGAMLNKLL 327

RESULT 2
ID Q9YEQ1 PRELIMINARY; PRT; 377 AA.
AC Q9YEQ1;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DE 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE 377AA LONG HYPOTHETICAL ACETAMIDASE.
GN APE0526.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1.
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000059; BAA79495.1; -
SQ SEQUENCE 377 AA; 41052 MW; 6FA6B2AE CRC32;

Query Match 8.2%; Score 197; DB 1; Length 377;
Best Local Similarity 34.2%; Pred. NO. 3.16e-18;
Matches 40; Conservative 27; Mismatches 47; Indels 3; Gaps 3;

Db 146 LGVGVASPKELLKEIKEREDLLKRGFVLPTPEGAVPREPVASEGLRTIPPENG 205
QY 131 LTAMLDLPEKVRMIKLDSEKVKYKSRHTLPYKPHIGTSLVSP-EIDSINSLTPDNHG 189
Db 206 NLDVRFHSPGKIYFPVEGA-LFSVGDHAYAOQDGVCGTATIMGTATATLRFQVI 261
QY 190 NMDVPDIPGSIYPLVRAPGRLE-IGDAHACQDGEICGTAVEFASITTIKVDLI 245

RESULT 3
ID O58755 PRELIMINARY; PRT; 389 AA.
AC O58755;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-JAN-1999 (TremBLrel. 09, Last annotation update)
DE 389AA LONG HYPOTHETICAL PROTEIN.
GN PH1041.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000004; BAA30139.1; -
SQ SEQUENCE 389 AA; 42005 MW; BAD3990F CRC32;

Query Match 6.7%; Score 163; DB 1; Length 389;

```

---

```

Best Local Similarity 38.7%; Pred. No. 1.19e-11;
Matches 24; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

Db 206 GHLDVDSVREGAIVAPYKVEGGYAGDAHAMEGDEGAGHTTDTAETTVTVVEIKGL 265
QY 189 GNMVDPIGPGSITYPLVRAPGRGLFIGDAHACQDGEICGTAVEFASITTIKVDLIKNN 248
Db 266 NL 267
QY 249 QL 250

RESULT 4
ID P78616 PRELIMINARY; PRT; 2091 AA.
AC P78616;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE FATTY ACID SYNTHASE, BETA SUBUNIT.
GN FASB.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Euscomycetes; Plecomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97121482.
RA BROWN D.W., ADAMS T.H., KELLER N.P.;
RT "Aspergillus has distinct fatty acid synthases for primary and
RT secondary metabolism.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14873-14877(1996).
DR EMBL; U75347; AAB41494.1; -
DR PFAM; PF00698; Acyl_transf; 1.
DR PFAM; PF01575; MaoC_like; 1.
SQ SEQUENCE 2091 AA; 232283 MW; 6B5D181B CRC32;

Query Match 4.6%; Score 110; DB 3; Length 2091;
Best Local Similarity 33.7%; Pred. NO. 1.64e-02;
Matches 28; Conservative 15; Mismatches 35; Indels 5; Gaps 4;

Db 467 IPVYGTGDDLRASIDANVVPALVRMTTHDP--VNWFEQTAFPNATHIVDFG-PGGISG 523
QY 120 IPHFGGLTGDTLAMLN-DLPEKVRMIKLDSEKVKYKSRHTLPYKPHIGTSLVSP 178
Db 524 LGVLTNRNKDGT-GVRVILAGSM 545
QY 179 INSLTPDNHGMDVDPDIGPSI 201

RESULT 5
ID Q22211 PRELIMINARY; PRT; 789 AA.
AC Q22211;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE SIMILARITY TO CYTOCHROME B.
GN T05B11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., COOPER J., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;

```





```

RL Plant Physiol. 108:861-861(1995).
DR EMBL; U17063; AAA86690.1; -.
DR MENDEL; 8699; Limdo; 1208:8699.
DR PFAM; PF00487; FA.desaturase; 1.
SQ SEQUENCE 436 AA; 50093 MW; 327D77FE CRC32;

Query Match      4.3%; Score 103; DB 10; Length 436;
Best Local Similarity 22.9%; Pred. No. 1.85e+01;
Matches 24; Conservative 29; Mismatches 42; Indels 10; Gaps 9;

Db    83 GSPPPE--KLADIRAAIPKHCWKQNRSSVYVDVVIVGLGAAAAVAANSW-AVWP-L-Y 139
     |::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY   229 GTAVEFASITIKVDLLTKN-W-QLSWRPMENAE-NIMSI-G-SARPLEDATRIAYRDLII 283
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    140 WQAQGTMF--WALFVLGHDCGGHGSFNHNKKNSVWGHLHHSSLV 182
     | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   284 WLVEDFEGEQDAYMILLSOGKYRLGNWDVPKYTVGAMLNKNLIV 328
           | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT       9
ID Q51454 PRELIMINARY; PRT; 473 AA.
AC Q51454;
DT 01-NOV-1996 (TrEMBLrel_01, Created)
DI 01-NOV-1996 (TrEMBLrel_01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel_12, Last annotation update)
DEF TRANSCRIPTIONAL ACTIVATOR.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK; MEDLINE: 96071911.
RX RITCHINGS B.W., ALMIRA E.C., LORY S., RAMPHAL R.; "Cloning and phenotypic characterization of fliC and flrA, new response regulators of Pseudomonas aeruginosa which regulate motility and adhesion to mucin." Infect. Immun. 63:4868-4876(1995).
RR EMBL; LA1213; AAA93530.1; -.
DR HSPB; PA1789; INTR.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
DR PROSITE; PS00676; SIGMA54_INTERACT_2; 1.
DR PROSITE; PS00688; SIGMA54_INTERACT_3; 1.
DR PFAM; PF00072; response_reg; 1.
DR PFAM; PF00158; sigma54; 1.
SQ SEQUENCE 473 AA; 51264 MW; BA4045A5 CRC32;

Query Match      4.3%; Score 105; DB 2; Length 473;
Best Local Similarity 30.8%; Pred. No. 9.38e+02;
Matches 20; Conservative 18; Mismatches 21; Indels 6; Gaps 6;

Db    128 DGPVALEPR-SROLLELARVARSDSTLVISGESGTGREVIANYIHQSOPRACKPFIAIN 186
     |:||:| ||:||:| :||| ::| :||:||||| :|:||||| |||:| :|
QY   62 EGAGINEQQIPIPSOLLAKMPF-LNPQNPTMVNG-AEKG-DVLAVYIESMLPGRVDPY-GI- 116
           ||| ||
Db    187 CAAIP 191
QY   117 CAMIP 121

RESULT       10
ID O45218 PRELIMINARY; PRT; 597 AA.
AC O45218;
DT 01-JUN-1998 (TrEMBLrel_06, Created)
DI 01-JUN-1998 (TrEMBLrel_06, Last sequence update)
DE DE AKLYLGIHDYOXYACETONEPHOSPATE SYNTHASE (EC 2.5.1.26)
DE (ALKYLGLYCERONE-PHOSPHATE SYNTHASE)
DE DE (ALKYLDHYDOXACYETONEPHOSPATE SYNTHASE).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoza; Secernentea; Rhabditiia; Rhabditiida; OC Rhabditiina; Rhabditoidea; Rhabditiidae; Pelodierinae; Caenorhabditis.
```

```
[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=CB1489 HIM-8(EI489); TISSUE=WHOLE ANIMAL;
RC      MEDLINE; 98113342.
RA      DE VET E.C.J.M., PRINSEN H.C.M.T., VAN DEN BOSCH H.;
RT      "Nucleotide sequence of a cDNA clone encoding a Caenorhabditis elegans
RT      homolog of mammalian alkyl-dihydroxyacetonephosphate synthase:
RT      evolutionary switching of peroxisomal targeting signals.";
RL      Biochem. Biophys. Res. Commun. 242:277-281(1998).
CC      -1- CATALYTIC ACTIVITY: 1-ACYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN
CC      ALCOHOL = 1-ALKYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN ACID ANION.
CR      ENBL; AJ002686; CAA05690.1; -.
DR      PRAM; PF01565; FAD_binding_4; 1.
KW      Transferrase
SQ      SEQUENCE 597 AA; 66559 MW; AFFB98EE CRC32;

Query Match          4.3%; Score 104; DB 5; Length 597;
Best Local Similarity 28.2%; Pred. No. 1.32e-01;
Matches 24; Conservative 25; Mismatches 29; Indels 7; Gaps 5;

Db    159 CAIIP-IGG--GTSVYNALDTPETERKAVISMMDALLDKILWIDRNLTCRAQAQIVGVS 215
QY    ||:||:|| || || | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY    117 CAMIPHEGGTGTDLTAMLNDPLPEKVRIKL-D-S--EKVYSKRHTLPYKPHIGTLSSV 173

Db    216 LERQLNKKGFTCGHEPDSIEFSTLG 240
QY    | : : : :| : : : :| : : : :|
QY    174 PEID-SINSITPDNHGGMVDPIG 197

RESULT 11
ID    O33765 PRELIMINARY; PRT; 1874 AA.
AC    O33765;
DT    01-JAN-1998 (TrEMBLrel. 05, Created)
DI    DT
DT    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DI    DT
DT    01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE    ICAL1 PROTEASE (EC 3.4.24.13).
GN    IGA.
OS    Streptococcus sanguis.
OC    Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC    Streptococcus.
RN    [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SK85;
RX      MEDLINE; 98084472.
RA      POULSEN K., REINHOLDT J., JESPERSGAARD C., BOYE K., BROWN T.A.,
RT      HAUGE M., KILLIAN M.;
RT      "A comprehensive genetic study of streptococcal immunoglobulin A1
RT      proteases: evidence for recombination within and between species.";
RL      Infect. Immun. 66:181-190(1998).
DR      ENBL; Y13461; CAA73858.1; -.
DR      PRAM; PF00746; Gram_pos_anchor; 1.
KW      Protease; Hydrolase; Metalloprotease.
SQ      SEQUENCE 1874 AA; 208478 MW; 4EF17F74 CRC32;

Query Match          4.3%; Score 104; DB 2; Length 1874;
Best Local Similarity 29.4%; Pred. No. 1.32e-01;
Matches 25; Conservative 21; Mismatches 33; Indels 6; Gaps 6;

Db    399 GGQSQAIVEPEQVSSLPYETGPQAGV-VE-PEQVAFLAEYIGPQAGSVVPEQVTPL- 455
QY    | : | : | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY    63 GAINSEDI-PSOLLKMP-FLNPQNGPIMVNGAEKDVLAVIESMLPRGVDPYGICAMI 120

Db    456 PETYGVOAGSVSPQATPLPEYTR 480
QY    | : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY    121 PHFGGL-TGTDLTAMLNDPLPEKVR 144

RESULT 12
ID    O17308 PRELIMINARY; PRT; 243 AA.
AC    O17308;
DT    01-JAN-1998 (TrEMBLrel. 05, Created)
DI    DT
DT    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DI    DT
DT    01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
```

| RESULT                | 14  |   |
|-----------------------|---|---|
| ID                    | Q29621  | PRELIMINARY; PRT: 4968 AA.              |
| AC                    | Q29621;   |   |
| DT                    | 01-NOV-1996   | (TREMBLrel. 01, Created)                |
| DT                    | 01-NOV-1996   | (TREMBLrel. 01, Last sequence update)   |
| DT                    | 01-NOV-1999   | (TREMBLrel. 12, Last annotation update) |
| DE                    | CARDIAC RYANODINE RECEPTOR.   |   |
| GN                    | RYR-2.  |   |
| OS                    | Oryctolagus cuniculus (Rabbit).   |   |
| OC                    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;           |   |
| OC                    | Eutheria; Lagomorpha; Leporidae; Oryctolagus.                           |   |
| [1]                   |   |   |
| SEQUENCE FROM N.A.    |   |   |
| RC                    | STRAIN=DOMESTIC RABBIT; TISSUE=HEART, BRAIN;                            |   |
| RC                    | MEDLINE: 910332060.   |   |
| RC                    | NAKAI J., IMAGAWA T., HAKAMAT Y., SHIGEKAWA M., TAKEISHIMA H., NUWA S.; |   |
| RT                    | "Primary structure and functional expression from cDNA of the cardiac   |   |
| RT                    | ryanodine receptor/calcium release channel.";                           |   |
| FEBS                  | Lett. 271:169-177(1990).  |   |
| [2]                   |   |   |
| SEQUENCE FROM N.A.    |   |   |
| RC                    | TISSUE=HEART, BRAIN;  |   |
| RC                    | MEDLINE: 93147068.  |   |
| RC                    | IMAGAWA T., NAKAI J., TAKEISHIMA H., NAKASAKI Y., SHIGEKAWA M.;         |   |
| RT                    | "Expression of Ca(2+)-induced Ca2+ release channel activity from        |   |
| RT                    | cardiac ryanodine receptor cDNA in Chinese hamster ovary cells.";       |   |
| J. Biochem.           | 112:508-513(1992).  |   |
| [3]                   |   |   |
| SEQUENCE FROM N.A.    |   |   |
| RC                    | STRAIN=DOMESTIC RABBIT; TISSUE=HEART, BRAIN;                            |   |
| RA                    | NAKAI J.;   |   |
| DR                    | Submitted (FEB-1996) to the EMBL/GenBank/DBDJ databases.                |   |
| DR                    | EMBL: U50465; AAA93465.1; -.  |   |
| DR                    | EMBL: U50466; AAA93469.1; -.  |   |
| DR                    | PFAM: PF01365; RYDR_ITPR; 1.  |   |
| DR                    | PFAM: PF00822; SPRY; 3.   |   |
| DR                    | PRINTS; PR00795; RYANODINER.  |   |
| SEQUENCE              | 4968 AA; 565081 MW; FDFE5E60 CRC32;                                     |   |
| Query Match           | 4.2%; Score 101; DB 6; Length 4968;                                     |   |
| Best Local Similarity | 26.6%; Pred. No. 3 63e-01;  |   |
| Matches               | 17; Conservative 19; Mismatches 24; Indels 4; Gaps                      |   |
| Db                    | 3385 LKBPTEAEELFRMAEVFIYWSKSHN--FKREEQFVYQNEINNMFLITDTKS-KMSK 3441      |   |
| QY                    | 135 LNDPLPEKVRMIKLDSEK-VYWSKRHTLPYKPHICTLSVSPIDSINSLTDPNHGGMVDV 193     |   |
| Db                    | 3442 AAVS 3445  |   |
| QY                    | 194 PDIG 197  |   |
| RESULT                | 15  |   |
| ID                    | Q30796  | PRELIMINARY; PRT: 281 AA.               |
| AC                    | Q30796;   |   |
| DT                    | 01-JAN-1998   | (TREMBLrel. 05, Created)                |
| DT                    | 01-JAN-1998   | (TREMBLrel. 05, Last sequence update)   |
| DT                    | 01-NOV-1998   | (TREMBLrel. 08, Last annotation update) |
| DE                    | MXAE.   |   |
| GN                    | MXAE.   |   |
| OS                    | Methylobacterium extorquens.  |   |
| OC                    | Bacteria; Proteobacteria; alpha subdivision; Methylobacterium.          |   |
| [1]                   |   |   |
| RC                    | STRAIN=AM1.   |   |
| RC                    | SEQUENCE FROM N.A.  |   |
| RC                    | MEDLINE: 98156126.  |   |
| RC                    | SPRINGER A.L., AUNAN A.J., LIDSTROM M.E.;                               |   |
| RT                    | "Sequence and characterization of mxAB, a response regulator involved   |   |
| RT                    | in regulation of methanol oxidation, and of mxAW, a methanol-regulated  |   |
| RT                    | gene in Methylobacterium extorquens AM1.";                              |   |
| RL                    | FEMS Microbiol. Lett. 160:119-124(1998).                                |   |
| DR                    | EMBL: AF017434; AAC46163.1; -.  |   |
| SEQUENCE              | 281 AA; 28758 MW; 5C08AFAA CRC32;                                       |   |

| Query Match           | 4.1%   | Score 99;  | DB 2;    | Length 281;                      |
|-----------------------|--------|--|----------|----------------------------------|
| Best Local Similarity | 25.9%; | Pred. No. 7  | 0.3e-01; |                                  |
| Matches               | 21;    | Conservative   | 23;      | Mismatches 33; Indels 4; Gaps 4; |
| Db                    | 127    | DGARWTRLATIPVG-TAPFALALSPDERLYVG-NVRS-NDLTVITDGTGLKAIAVPAGA  | 183      |                                  |
| QY                    | 141    | EKVRMIKLDSEKVNYSKRHTLPVKPHIGLTVSYSPEDTSINSLTPDHGNGMDVPDIPGGS | 200      |                                  |
| Db                    | 184    | MPYGVSYSPDGRVFTVNOHA   | 204      |                                  |
| QY                    | 201    | ITYPL-VRAPGGRLFIGDAHA  | 220      |                                  |

Search completed: Thu Aug 24 18:53:01 2000  
Job time : 69 secs.